# GENEROUTE: SPATIAL GENE EXPRESSION PREDICTION FROM H&E HISTOLOGY WITH MOLECULAR PRIORS AND CANCER CONTEXT

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Predicting spatial gene expression from routine H&E histology could make molecular tissue maps widely accessible, but fixed output panels restrict many models to genes observed during training. We introduce GeneRoute, a gene-conditioned framework that uses molecular knowledge to guide multiscale image interpretation. Sequence and functional priors define shared gene representations, while cancer-specific dependencies and molecular programs adapt them to tissue context. These queries retrieve histological features at three scales, and a shared scalar decoder predicts expression without gene-specific output heads. Across five cancers, GeneRoute leads all five macro-averaged metrics on cancer-related genes, improving gene-wise Pearson correlation by **25%** over the strongest comparator. For **200 genes** excluded from expression supervision and model selection, it also achieves the highest macro-averaged gene-wise Pearson and Spearman correlations. On the six-gene immune-receptor subset, GeneRoute achieves positive mean gene-wise correlations across all five cancers, with predicted IGHG1 enrichment partly overlapping outlined immune infiltrates in a colorectal section.

## 1 INTRODUCTION

Spatial transcriptomics (ST) measures gene expression while preserving tissue organization, enabling the study of molecular heterogeneity in its spatial context (Ståhl et al., 2016). Its application remains constrained by assay cost, experimental requirements, and platform-dependent gene coverage. Hematoxylin-and-eosin-stained (H&E) histology is routinely collected and records cellular morphology and tissue architecture. This combination has motivated models that predict spatial expression from histological images (He et al., 2020; Fu et al., 2025). Most such models, however, predict a predefined output vector whose coordinates correspond to genes observed during training. Extending their predictions to another gene typically requires changing the output space and obtaining additional expression supervision.

Histology-to-expression methods have improved through richer visual and spatial representations. Direct image-to-expression models were followed by approaches that incorporate neighboring spots, graph structure, or multiple fields of view (He et al., 2020; Chen et al., 2024; Zeng et al., 2022; Xiao et al., 2024; Chung et al., 2024). More recent methods explore generative objectives and transferable image–expression representations (Huang et al., 2025a;b; Wu et al., 2026). Broader transcriptome coverage, however, does not by itself resolve how gene-specific knowledge should guide multiscale morphology. Gene-aware approaches, including GeneQuery, SGN, and DeepSpot-M, address this question by treating gene information as an input rather than only as an output index (Xiong et al., 2024; Yang et al., 2024; Nonchev et al., 2026).

Using gene knowledge to interpret morphology requires separating shared gene identity from cancer-specific context. Sequence and functional annotations describe relatively stable aspects of a gene, whereas its dependencies and associated molecular programs can differ across cancer contexts (Pacini et al., 2024; Arafeh et al., 2025). Relevant histological information can also occur at different scales, from local cellular organization to broader tissue structure. A useful model should therefore preserve shared gene information, adapt it to the biological context, and use the adapted representation to select relevant morphological features.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We introduce GENEROUTE, a framework that combines these three levels of information. Frozen protein-sequence, RNA-sequence, and functional-text embeddings are combined with general gene relationships to construct a shared representation (Lin et al., 2023; Chen et al., 2022; Chen & Zou, 2025; Szklarczyk et al., 2023). Cancer-specific dependency graphs adapt this representation to the requested cancer context (Pacini et al., 2024; Arafeh et al., 2025), while source-separated molecular program banks provide context-dependent gene information, including perturbation-morphology profiles from JUMP (Chandrasekaran et al., 2024). The adapted gene query attends to H&E features at three spatial scales, and a gene- and location-dependent router combines the resulting image readouts. A shared decoder predicts one expression value per gene and spatial location. This formulation supports queries for genes excluded from spatial-expression supervision, provided their molecular priors are available.

We evaluate GENEROUTE on five cancer contexts using common benchmark data partitions. The evaluation combines quantitative benchmarks with two spatial analyses: EPCAM localization in four epithelial cancers and immune-gene prediction after exclusion from expression supervision. We also examine breast-cancer markers and cell-abundance maps as supplementary tissue context. Our contributions are threefold:

- We construct gene representations that combine complementary molecular priors while explicitly accounting for missing sources and separating shared gene information from cancer-specific context.
- We introduce cancer-adapted gene queries that guide multiscale histology readout and a shared, context-modulated decoder that avoids gene-specific output heads.
- We evaluate supervised and supervision-unseen gene prediction across five cancer contexts, with quantitative benchmarks, epithelial and immune spatial examples, and component ablations.

## 2 RELATED WORK

Histology-to-expression methods use H&E morphology to estimate spatial expression. Early work established local and whole-slide prediction, followed by attention- and graph-based models that incorporate neighboring spots (Schmauch et al., 2020; He et al., 2020; Pang et al., 2021; Zeng et al., 2022; Jia et al., 2024; Xiao et al., 2024). HE2Gene and BrSTNet further develop image-based prediction, while multi-scale methods such as TRIPLEX and M2OST combine local and broader image evidence (Chen et al., 2024; Rahaman et al., 2023; Chung et al., 2024; Wang et al., 2025b). Contrastive and generative objectives extend representation learning (Xie et al., 2023; Min et al., 2024; Huang et al., 2025a;b). These advances improve visual and spatial encoding, but most models still predict a fixed panel learned from expression supervision. Benchmarks and shared resources support comparisons across datasets and modeling paradigms (Jaume et al., 2024; Wang et al., 2025a).

Gene-aware models make gene identity an input to the prediction process. GeneQuery introduces gene queries, SGN uses functional and phenotype descriptions to support genes without spatial-expression supervision, and DeepSpot-M combines gene queries with a hypernetwork for transcriptome-scale prediction (Xiong et al., 2024; Yang et al., 2024; Nonchev et al., 2026). AGP-Net and CoMTIP also use gene descriptions or joint gene–expression–image modeling (Yang et al., 2025; Qian et al., 2025). These studies establish gene queries as a route to prediction without expression supervision for every requested gene.

GENEROUTE builds on gene-conditioned prediction by separating shared molecular representations from cancer-specific adaptation. The adapted queries guide multiscale histology readout before shared scalar decoding. This combines gene identity, cancer context, and image-scale selection in one prediction pathway. See Section Q for extended related work and molecular-resource descriptions.

## 3 METHOD

Figure 1 summarizes GENEROUTE. For each requested gene, frozen molecular priors are fused into a shared query, adapted with cancer-context resources, and used to read registered H&E features at three scales. A location- and gene-dependent router combines the image readouts before a shared

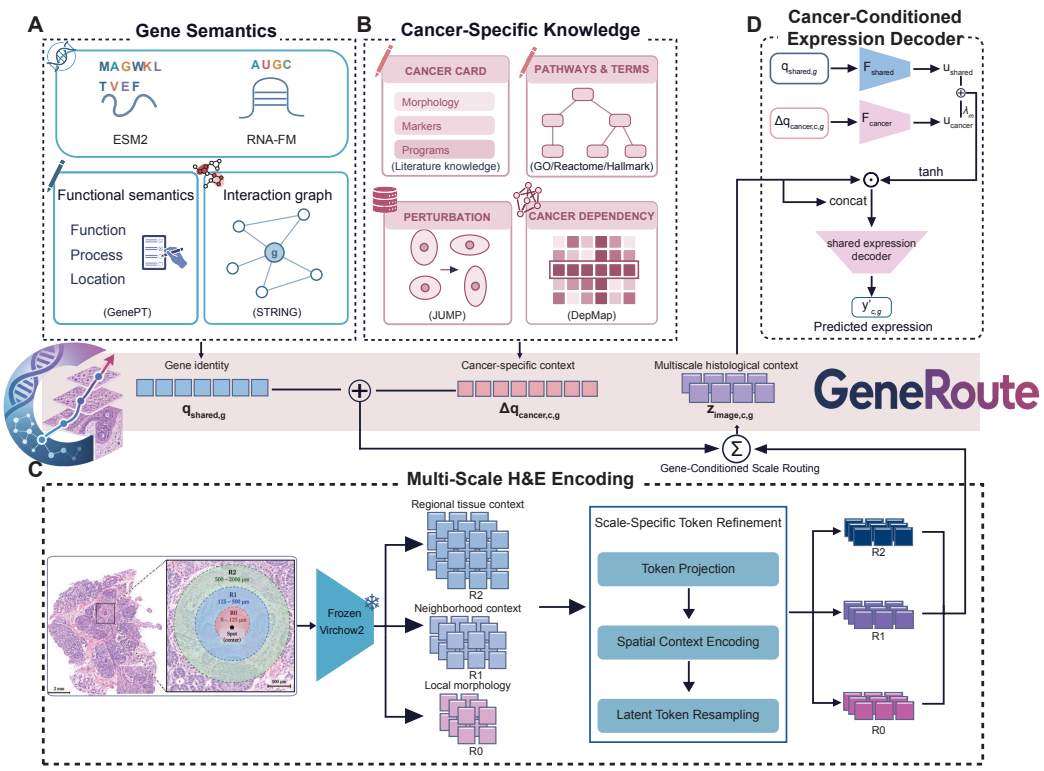


Figure 1: **GeneRoute architecture.** Molecular priors form a shared gene query, cancer-context resources adapt it, and the resulting query reads registered H&E features at three scales. A learned router and shared decoder produce one expression value per requested gene and location.

scalar decoder predicts expression. The exact equations and implementation details are given in Section A.

### 3.1 Gene-conditioned query construction

Let $b$ denote a spatial location, $c$ its cancer context, and $g$ the requested gene. The model predicts one value per gene–location pair, $\widehat{y}_{b,c,g} = f_\theta(I_b, \mathcal{K}_g, \mathcal{P}_c)$, using shared parameters rather than a fixed output head. A gene is queryable when it maps to the identity registry and has at least one molecular embedding; evaluating it additionally requires a measured target. Source availability and expression availability are masked separately, and missing expression is never treated as zero.

Frozen ESM2, RNA-FM, and GenePT vectors are projected to $D = 512$ and availability-fused with a shared gene-type embedding:

$$\begin{aligned} e_{g,m} &= \mathrm{LN}(P_m(\mathrm{embedding}_{g,m})), \\ q_g^0 &= \mathrm{LN}\left(\sum_m \alpha_{g,m} e_{g,m} + e_{\mathrm{type},g}\right), \\ \alpha_{g,m} &= \frac{A_{g,m} e^{a_m(e_{g,m})}}{\sum_{m'} A_{g,m'} e^{a_{m'}(e_{g,m'})}}. \end{aligned} \tag{1}$$

Here $A_{g,m}$ records source availability and $a_m$ is a learned source score. The shared query is refined by the sparse STRING residual

$$q_{\mathrm{shared},g} = \mathrm{LN}\left(q_g^0 + \gamma_{\mathrm{STRING}} \sum_{h \in N_{\mathrm{STRING}}(g)} w_{g,h} W_{\mathrm{STRING}} q_h^0\right). \tag{2}$$

Cancer context then selects a DepMap graph and six source-separated program banks (Cancer Card, Hallmark, Gene Ontology, Reactome, JUMP ORF, and JUMP CRISPR):

$$\begin{aligned} q'_{c,g} &= \mathrm{LN}(q_{\mathrm{shared},g} + \lambda_{\mathrm{ctx}} \Delta q_{\mathrm{cancer},c,g}) \,, \\ \Delta q_{\mathrm{cancer},c,g} &= \gamma_{\mathrm{DepMap}} r_{\mathrm{DepMap},c,g} \\ &\quad + \sum_m \eta_m v_{m,c} \,\mathrm{sigmoid}(G_m(q_{\mathrm{shared},g})) \\ &\quad \times \mathrm{CrossAttention}(q_{\mathrm{shared},g}, K_{m,c}, V_{m,c}). \end{aligned} \tag{3}$$

Here $v_{m,c}$ masks invalid banks and $G_m$ is a learned gate. The cancer label chooses the context resources but is not passed directly to the decoder. Missing graphs or banks contribute zero residuals; priors encode external knowledge or perturbation-derived profiles, not measured activity in the target section. The source masks, graph refinement, and bank validity rules are detailed in Sections A, B, D and E.

### 3.2 Multiscale image readout and shared decoding

Each target spot is represented by three registered Virchow2 contexts, R0, R1, and R2, with anchor capacities 7, 32, and 64. Scale adapters, positional encodings, shared spatial attention, and a learned 16-query resampler produce $X_{\mathrm{image}} \in \mathbb{R}^{B\times 3\times 16\times 512}$. Image keys and values contain H&E features only. The adapted gene query reads each scale and a learned router combines them:

$$\begin{aligned} z_{b,c,g,s} &= \mathrm{CrossAttention}(q'_{c,g}, K_{b,s}, V_{b,s}), \quad r_{c,g} = \tanh(W_{\mathrm{route}} q'_{c,g}), \\ w_{b,c,g,s} &= \mathrm{softmax}_s \, F_{\mathrm{scale}}([z_{b,c,g,s}, z_{b,c,g,s} \odot r_{c,g}]), \quad z_{\mathrm{image},b,c,g} = \sum_s w_{b,c,g,s} z_{b,c,g,s}. \end{aligned} \tag{4}$$

The attention output has no additive molecular-query residual, so the readout remains image-derived. Shared and cancer-context conditioner networks modulate its channels while preserving the unmodulated path:

$$\begin{aligned} m_{c,g} &= \tanh(F_{\mathrm{shared}}(q_{\mathrm{shared},g}) + \lambda_m F_{\mathrm{cancer}}(\Delta q_{\mathrm{cancer},c,g})) \,, \\ \widehat{y}_{b,c,g} &= \mathrm{SharedDecoder}([z_{\mathrm{image},b,c,g}, z_{\mathrm{image},b,c,g} \odot m_{c,g}]) \,. \end{aligned} \tag{5}$$

The same scalar decoder is used for every gene and cancer context. Exact tensor dimensions and conditioner architecture are given in Section H.

### 3.3 Training and complexity

GeneRoute targets use library-size normalization followed by $\log(1+\cdot)$ transformation. Stage 1 uses masked Huber loss; stage 2 uses masked half-MSE plus $0.1$ within-section gene-wise PCC loss, with at least eight valid spots and at most 128 spots per gene–section group. Correlations never pool unrelated sections or cancer contexts. Batches are cancer-balanced, genes are processed in chunks of 256, and the epoch-level sampling schedule contains a $0.8$ uniform component sampled once per epoch and a $0.2$ priority component trained once per epoch. The selected run used AdamW with base learning rate $10^{-4}$, weight decay 0, global-norm gradient clipping at 1.0, and bfloat16 mixed precision. This schedule maintains broad gene coverage while including priority genes.

Virchow2, molecular encoders, and precomputed program embeddings are frozen; the remaining modules are trained jointly. With $S = 3$ scales, $L = 16$ latents per scale, latent width $D$, and $G$ requested genes, gene–image cross-attention costs approximately $\mathcal{O}(BGSLD)$ and chunking bounds peak memory by the gene chunk. The target, loss, sampling, and run details are given in Sections A, I, K and O.

## 4 Experimental Setup

**Cohorts and common data splits.** We evaluate colorectal, prostate, kidney RCC, brain/glioma, and breast cancer using publicly available H&E-paired Visium cohorts, primarily HEST-1K and supplemented by other public datasets (Jaume et al., 2024). The training partition contains 108 sections and 318,458 spots, the validation partition contains 23 sections and 82,484 spots, and the common test pool contains 33 sections and 94,049 spots (Table S4). All three partitions are disjoint at both the patient and specimen levels. Missing targets are masked.

**Expression targets and reporting panels.** The source axis has 6,100 genes, of which 6,088 map to the query registry. We report training-set-defined high-variance (1,000), low-variance (500), HVG (1,000), and cancer-related (100 nominal; 60–68 supported per cancer) panels. The same requested and scored panels are used for every method. All four panels exclude the permanent held-out set, $\mathcal{U}_{200}$, but can overlap each other (Section J). GeneRoute and the common-target baselines use the transformation in Eq. (30); GeneQuery uses its native section-wise target in Eq. (35), with missing measurements omitted from its extrema and loss. Reference and held-out information access are summarized in Table A7.

**Baselines.** We compare with HE2Gene, ST-Net, HisToGene, Hist2ST, THItoGene, BrSTNet, TCGN, STFlow, TissueFormer, GeneQuery, SGN, and DeepSpot-M. All methods use the same train, validation, and test partitions and the same scored gene panels. GeneQuery uses its native target, TissueFormer uses in-context retrieval, and SGN is trained directly on the common target for supervised panels and uses training-reference calibration only for $\mathcal{U}_{200}$. Target transformations and permitted reference inputs are detailed in Section L and Table A7.

**Metrics and aggregation.** We report spot-wise and gene-wise Pearson and Spearman correlations and RMSE. Correlations are computed within sections, averaged equally within cancer, and then averaged equally across the five cancer means; RMSE is averaged from section-level errors. Undefined correlations are omitted from section means, never replaced by zero, and accompanied by constant-gene and constant-spot counts; common requested panels can therefore have different defined subsets after constant masking. A macro score is unavailable when any cancer lacks a defined score. Standard deviations describe cancer means or test-section variation, not confidence intervals or seed variation. GeneQuery's native-target gene-wise values are shown for reference and excluded from common-target ranking; spot-wise values and RMSE are omitted. Full definitions, thresholds, and target conventions are given in Section L.1.

**Training and model selection.** The main GENEROUTE results use a pooled model and the two-stage objective in Section A.6. For GeneRoute, the permanent held-out set is excluded from both expression objectives from the start of stage 1, including the stage-1 checkpoint used for continuation and stage 2; it is also excluded from validation scores, early-stopping decisions, and hyperparameter selection. All benchmark methods use a maximum of 100 training epochs and validation-based early stopping. Component ablations use a fixed budget of 20 epochs of stage 1. Checkpoint and baseline procedures are in Sections K and L.

**Spatial visualization.** EPCAM maps use method-specific color limits, while their PCC values use expression predictions rather than rendered colors. Supplementary breast-cancer maps use a fixed per-gene range for common-target methods; GeneQuery retains its native section-wise target.

# 5 RESULTS

## 5.1 PREDICTION PERFORMANCE ON SUPERVISED GENES

GENEROUTE leads the five-cancer macro average on all five cancer-related-panel metrics (Table 1): spot PCC 0.7392, gene PCC 0.1249, spot/gene Spearman 0.6060/0.1267, and RMSE 0.6786, versus 0.7073/0.0999 spot/gene PCC for STFlow. It has the highest spot PCC in all five contexts and the highest gene PCC in four epithelial cancers; brain/glioma favors STFlow (0.0623 versus 0.0368).

Variation across cancers and sections is shown by the reported standard deviations, and rankings change by panel (Supplementary Tables S1 and S2). STFlow leads the high-variance panel and leads the spot-wise metrics and RMSE on the other broad panels. GENEROUTE leads gene-wise PCC/Spearman on low-variance (0.0372/0.0773) and gene-wise Spearman on HVG (0.0938), so its advantage depends on the panel and metric.

Table 1: **Prediction performance on cancer-related genes.** Five-cancer macro averages for five metrics are shown for all methods. The nominal panel has 100 genes, with the same 60–68 supported genes per cancer scored across methods; per-cancer spot-wise and gene-wise PCC values are reported in Supplementary Table A8. Values are means with standard deviations; aggregation is described in the notes.

| | Five-cancer macro average | | | | |
|---|---|---|---|---|---|
| **Method** | $\text{PCC}_{\text{spot}} \uparrow$ | $\rho_{\text{spot}} \uparrow$ | $\text{PCC}_{\text{gene}} \uparrow$ | $\rho_{\text{gene}} \uparrow$ | RMSE$\downarrow$ |
| HE2Gene | 0.5334 (0.138) | 0.4993 (0.108) | 0.0532 (0.036) | 0.0544 (0.037) | 0.8788 (0.085) |
| ST-Net | 0.5051 (0.197) | 0.4612 (0.169) | 0.0225 (0.029) | 0.0262 (0.031) | 0.9006 (0.171) |
| HisToGene | 0.5965 (0.107) | 0.5489 (0.116) | 0.0358 (0.028) | 0.0431 (0.032) | 0.8246 (0.076) |
| Hist2ST | 0.5560 (0.134) | 0.5194 (0.120) | 0.0004 (0.004) | -0.0001 (0.003) | 0.8474 (0.078) |
| THItoGene | 0.5449 (0.135) | 0.5086 (0.123) | 0.0002 (0.003) | 0.0004 (0.004) | 0.8530 (0.085) |
| BrSTNet | 0.5199 (0.168) | 0.4822 (0.153) | 0.0129 (0.024) | 0.0087 (0.030) | 0.8611 (0.108) |
| DeepSpot-M | 0.5298 (0.165) | 0.5047 (0.124) | 0.0311 (0.021) | 0.0438 (0.024) | 0.8801 (0.082) |
| STFlow | <u>0.7073</u> (0.083) | <u>0.5933</u> (0.107) | <u>0.0999</u> (0.030) | <u>0.1035</u> (0.029) | <u>0.6996</u> (0.058) |
| TCGN | 0.4972 (0.182) | 0.4442 (0.154) | 0.0413 (0.032) | 0.0517 (0.038) | 0.9215 (0.124) |
| GeneQuery[§] | – | – | 0.0319 (0.030) | 0.0380 (0.043) | – |
| SGN | 0.4081 (0.175) | 0.3185 (0.070) | 0.0148 (0.025) | 0.0225 (0.028) | 0.9361 (0.092) |
| TissueFormer[†] | 0.5731 (0.159) | 0.4732 (0.178) | 0.0554 (0.035) | 0.0634 (0.036) | 0.8457 (0.083) |
| GeneRoute | **0.7392** (0.097) | **0.6060** (0.102) | **0.1249** (0.051) | **0.1267** (0.054) | **0.6786** (0.071) |

*Notes.* Parentheses give sample SD across five cancer means ($n = 5$). PCC and $\rho$ denote Pearson and Spearman correlations. Bold/underline mark best/second on the common target. § GeneQuery uses its native target, is excluded from ranking, and has no common-target spot-wise metrics or RMSE. † TissueFormer uses training-reference retrieval; supervised SGN outputs the common target (Table A7).

### 5.2 EPCAM LOCALIZATION ACROSS EPITHELIAL CANCERS

EPCAM is an established epithelial marker that is also expressed in nonmalignant epithelium (Went et al., 2006). We use it to examine the same gene across breast, colorectal, kidney RCC, and prostate contexts; this is an epithelial-associated rather than tumor-specific analysis.

In one representative test section per context (Fig. 2), GENEROUTE recovers major EPCAM-enriched regions and has the highest section-level PCC among the compared methods. Method-specific color limits expose spatial structure despite different ranges; neither these maps nor PCC establishes accurate expression amplitudes.

Supplementary Figure S1 adds the remaining baselines for colorectal and prostate. TENX53 ESR1, ERBB2, AR, and inferred cell-abundance maps provide additional common-scale context (Section M.2).

### 5.3 SPATIAL PREDICTION OF GENES EXCLUDED FROM EXPRESSION SUPERVISION

**Held-out gene evaluation.** The panel contains 138 genes annotated `protein_coding` and 62 genes with other GENCODE biotypes, selected by prior availability (Section J). All GeneRoute expression training and model selection exclude these genes. On the common 33 sections and 94,049 spots, GENEROUTE leads macro gene-PCC and gene-Spearman across all 200 genes (0.0273/0.0515; Table 2). Per-cancer results appear in Supplementary Table S3.

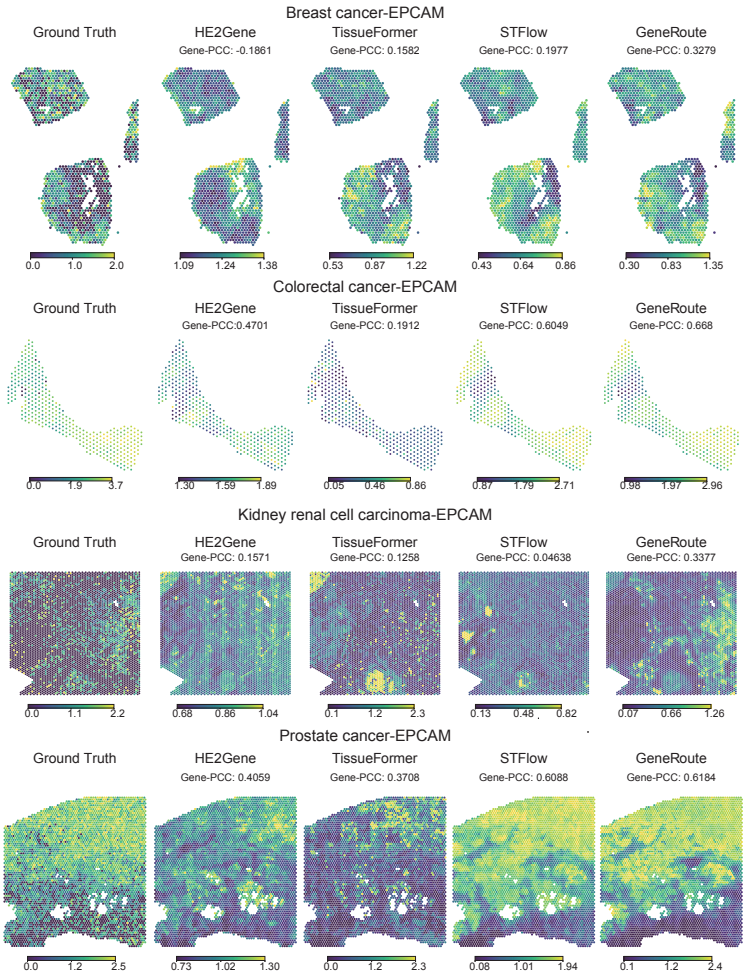


Figure 2: **EPCAM localization across four epithelial cancers.** Sections are CID4465 (breast), ZEN38 (colorectal), INT20 (kidney RCC), and TENX40 (prostate). GENEROUTE achieves the highest within-section PCC among the compared methods. Maps use separate color limits.

**Immune-associated spatial expression.** Motivated by the spatial organization of tumor-infiltrating lymphocytes (Cabrita et al., 2020), we examine six immune-receptor constant-region genes in $\mathcal{U}_{200}$: IGHG1, IGHG3, IGLC1, TRGC1, TRBC1, and TRBC2. GENEROUTE has the highest macro gene-PCC/Spearman (0.1000/0.1007 versus 0.0541/0.0606 for reference-assisted TissueFormer), with positive gene-PCC in all cancers (0.0319–0.1592); TissueFormer has the lowest subset RMSE (Table 2).

**Spatial correspondence with immune infiltrates.** On colorectal section ZEN46, GENEROUTE gives the highest IGHG1 PCC among the compared methods (0.285) and shows partial spatial overlap with the outlined immune-infiltrated regions (Fig. 3).

Table 2: **Held-out gene prediction.** Five-cancer macro results on 33 test sections for (A) all 200 genes and (B) the six-gene immune subset.

| Method | $\text{PCC}_{\text{gene}}$ ↑ | $\rho_{\text{gene}}$ ↑ | RMSE↓ |
|---|---|---|---|
| **(A) All 200 held-out genes** | | | |
| DeepSpot-M | 0.0167 (0.008) | <u>0.0402</u> (0.020) | <u>0.4410</u> (0.130) |
| TissueFormer† | <u>0.0184</u> (0.008) | 0.0304 (0.014) | **0.3768** (0.044) |
| SGN‡ | 0.0103 (0.011) | 0.0269 (0.023) | 0.4448 (0.129) |
| GeneQuery§ | 0.0197 (0.017) | 0.0217 (0.022) | – |
| GENEROUTE | **0.0273** (0.010) | **0.0515** (0.023) | 0.4535 (0.114) |
| **(B) Six immune-receptor genes** | | | |
| DeepSpot-M | -0.0752 (0.061) | -0.0696 (0.061) | 1.0682 (0.714) |
| TissueFormer† | <u>0.0541</u> (0.045) | <u>0.0606</u> (0.053) | **0.9184** (0.188) |
| SGN‡ | -0.0891 (0.077) | -0.0773 (0.077) | 1.0705 (0.631) |
| GeneQuery§ | 0.0116 (0.064) | 0.0194 (0.079) | – |
| GENEROUTE | **0.1000** (0.046) | **0.1007** (0.043) | <u>0.9636</u> (0.591) |

Means (SD) across cancers; bold/underline: best/second on the common target. † training-reference retrieval; ‡ training-reference calibration; § native target, excluded from ranking. Reference access: Table A7.

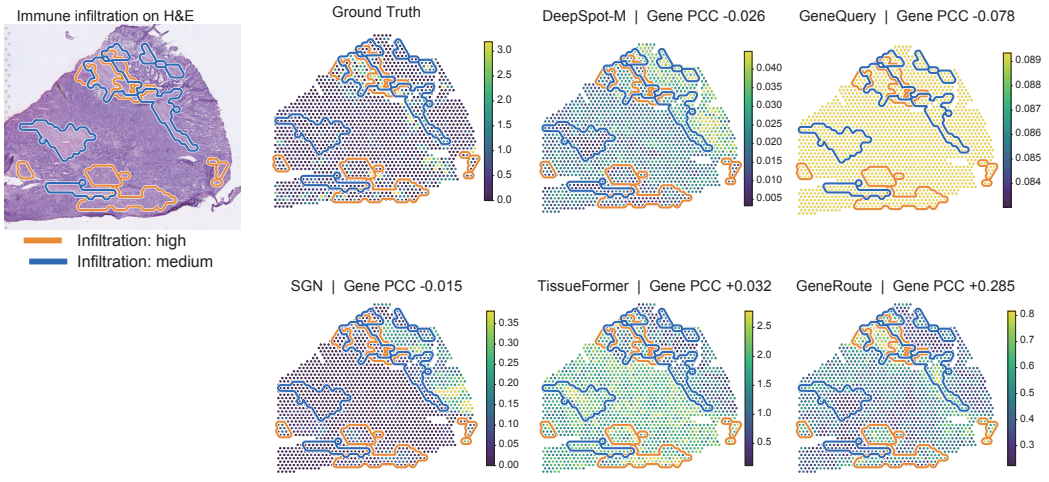


Figure 3: **Held-out immune gene IGHG1 on colorectal section ZEN46.** The infiltration contours are overlaid on measured expression and predictions from the five methods in Table 2; titles report within-section gene-PCC and maps use separate limits. GENEROUTE has the highest PCC among the compared methods (0.285), with partial overlap of high-expression regions and annotated infiltrates. The overlay marks infiltration rather than gene expression, cell identity, or confirmed TLS.

Table 3: **Component ablation.** Section-averaged gene-wise PCC on four panels for the same 20-section subset (four sections per cancer) randomly selected from the 33-section common test pool. Δ vs Full is the percentage difference in the four-panel mean relative to the matched complete model trained for 20 stage-1 epochs.

| **Condition** | High-var | Low-var | HVG | Cancer | Mean | Δ vs Full |
|---|---|---|---|---|---|---|
| Full model (20 epochs) | 0.1039 | 0.0364 | 0.0592 | 0.1020 | 0.0754 | 0.0 |
| *Component perturbations* | | | | | | |
| w/o priority sampling | 0.0723 | 0.0292 | 0.0402 | 0.0472 | 0.0472 | −37.4 |
| R0 only (w/o multi-scale) | 0.0851 | 0.0237 | 0.0498 | 0.0817 | 0.0601 | −20.3 |
| Random sequence embeddings | 0.0830 | 0.0251 | 0.0527 | 0.0841 | 0.0612 | −18.8 |
| w/o all priors | 0.0928 | 0.0238 | 0.0500 | 0.0917 | 0.0646 | −14.4 |
| w/o DepMap | 0.0910 | 0.0272 | 0.0552 | 0.0853 | 0.0647 | −14.2 |
| w/o program prior KV | 0.0960 | 0.0306 | 0.0586 | 0.0884 | 0.0684 | −9.3 |
| w/o JUMP programs | 0.0971 | 0.0298 | 0.0582 | 0.0880 | 0.0683 | −9.4 |
| w/o decoder cancer modulation | 0.0969 | 0.0331 | 0.0558 | 0.0882 | 0.0685 | −9.2 |
| w/o enrichment programs | 0.1009 | 0.0330 | 0.0569 | 0.0900 | 0.0702 | −6.9 |
| w/o STRING | 0.1008 | 0.0295 | 0.0559 | 0.0952 | 0.0704 | −6.7 |
| w/o GenePT | 0.1048 | 0.0306 | 0.0561 | 0.0996 | 0.0727 | −3.5 |
| w/o cancer card | 0.1023 | 0.0326 | 0.0593 | 0.0937 | 0.0720 | −4.5 |

*Notes.* The complete model and all variants use the same training budget of 20 epochs of stage 1, 20 test sections, four reporting panels, and masked gene-PCC definition. Source removals and cohort scoring are detailed in Section N.

### 5.4 COMPONENT ABLATION

We retrain the complete model and component variants for 20 epochs of stage 1 on the common training partition (Table 3). We randomly selected four test sections per cancer from the 33-section common test pool, for 20 sections in total; the same sections and scored panels were used for the complete model and every perturbation. Relative to the matched complete model (mean gene-PCC $= 0.0754$), mean gene-PCC reductions are $37.4\%$ without priority sampling, $20.3\%$ with R0 only, $18.8\%$ with random sequence embeddings, $14.2\%$ without DepMap, and $14.4\%$ without all priors (Section N).

## 6 Discussion and Limitations

**Biological interpretation.** GeneRoute captures morphology–expression associations at the resolution of Visium spots, which aggregate multiple cells. EPCAM maps and held-out immune-gene predictions connect these associations to epithelial and immune tissue organization. Cell-abundance estimates and infiltration contours provide complementary cellular and histological context for these spatial patterns.

**Evaluation and source coverage.** The patient- and specimen-disjoint evaluation characterizes prediction for new sections and supervision-unseen genes within five cancer contexts. Gene queryability depends on external-prior coverage, and performance varies across panels and cancers. Transfer to new cancer contexts or acquisition settings lies outside this evaluation. Gene chunking bounds peak attention memory, while inference cost grows with the requested panel size.

**Interpreting the performance differences.** Priority sampling produces the largest ablation effect, identifying the allocation of training across genes as a major contributor to performance. The stronger gains on cancer-related genes are consistent with this training emphasis, although the present ablations do not separate its interaction with each molecular prior. For $\mathcal{U}_{200}$, higher correlations but higher RMSE than TissueFormer distinguish spatial-pattern transfer from expression-amplitude calibration. The $\mathrm{R0}$-only comparison supports the value of additional tissue context; separating learned routing from simple multiscale fusion remains a useful controlled comparison.

**Future directions.** Further evaluation should test whether gene-dependent scale weights align with independently annotated tissue structures and remain stable across sections and acquisition settings. Controlled masking of image regions and removal of prior sources can assess the evidence supporting individual predictions. Extending the framework to genetic or drug responses would require paired perturbation–response supervision.

## 7 Conclusion

We introduced GeneRoute, which combines molecular priors, cancer context, and multiscale histology through a shared decoder. It leads the cancer-related panel on five macro metrics, recovers EPCAM patterns in four illustrated contexts, and leads macro gene-wise correlations for 200 supervision-unseen genes and a six-gene immune subset. These results support gene-conditioned prediction beyond the supervised panel within the evaluated cancer contexts.

## AI USE STATEMENT

Generative AI tools were used to assist with language editing, manuscript organization, manuscript critique, reference verification, and clarification of the evaluation protocol during revision. The authors take responsibility for the scientific content, reported results, and references.

## ETHICS STATEMENT

This work predicts spatial expression for research use. Outputs are computational estimates, not experimental measurements or clinical molecular tests. The analyses are secondary analyses of publicly available data, and no human specimens were newly collected for this study. Potential risks include reproducing biases in the training cohorts and external knowledge sources and over-interpreting predictions for diagnosis or treatment selection. Reuse of data and model outputs is subject to the original resource licenses and applicable privacy and institutional requirements.

## REPRODUCIBILITY STATEMENT

The manuscript and appendix describe the architecture, training objectives, expression preprocessing, source masks, and evaluation procedure. The selected pooled GeneRoute run used AdamW with base learning rate $10^{-4}$, weight decay 0, global-norm gradient clipping at 1.0, and bfloat16 mixed precision. The data composition, training configuration, panel definitions, and supplementary gene lists are indexed in Section P.

## REFERENCES

10x Genomics. Human breast cancer: Visium fresh frozen, whole transcriptome. 10x Genomics dataset, July 2022. URL https://www.10xgenomics.com/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard. Published July 1, 2022.

Rand Arafeh, Tsukasa Shibue, Joshua M. Dempster, William C. Hahn, and Francisca Vazquez. The present and future of the cancer dependency map. *Nature Reviews Cancer*, 25(1):59–73, 2025. doi: 10.1038/s41568-024-00763-x. URL https://doi.org/10.1038/s41568-024-00763-x.

Rita Cabrita, Martin Lauss, Adriana Sanna, Marco Donia, Mathilde Skaarup Larsen, Shamik Mitra, Iva Johansson, Bengt Phung, Katja Harbst, Johan Vallon-Christersson, Alison van Schoiack, Kristina Lövgren, Sarah Warren, Karin Jirström, Håkan Olsson, Kristian Pietras, Christian Ingvar, Karolin Isaksson, Dirk Schadendorf, Henrik Schmidt, Lars Bastholt, Ana Carneiro, Jennifer A. Wargo, Inge Marie Svane, and Göran Jönsson. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*, 577:561–565, 2020. doi: 10.1038/s41586-019-1914-8.

Loïc Chadoutaud, Marvin Lerousseau, Daniel Herrero-Saboya, Julian Ostermaier, Jacqueline Fontugne, Emmanuel Barillot, and Thomas Walter. sCellST predicts single-cell gene expression from H&E images. *Nature Communications*, 17(1):1194, 2026. doi: 10.1038/s41467-025-67965-1. URL https://doi.org/10.1038/s41467-025-67965-1.

Srinivas Niranj Chandrasekaran, Beth A. Cimini, Amy Goodale, Lisa Miller, Maria Kost-Alimova, Nasim Jamali, John G. Doench, Briana Fritchman, Adam Skepner, Michelle Melanson, Alexandr A. Kalinin, John Arevalo, Marzieh Haghighi, Juan C. Caicedo, Daniel Kuhn, Desiree Hernandez, James Berstler, Hamdah Shafqat-Abbasi, David E. Root, Susanne E. Swalley, Sakshi Garg, Shantanu Singh, and Anne E. Carpenter. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nature Methods*, 21(6):1114–1121, 2024. doi: 10.1038/s41592-024-02241-6. URL https://doi.org/10.1038/s41592-024-02241-6.

Jiayang Chen, Zhihang Hu, Siqi Sun, Qingxiong Tan, Yixuan Wang, Qinze Yu, Licheng Zong, Liang Hong, Jin Xiao, Tao Shen, Irwin King, and Yu Li. Interpretable RNA foundation model from

unannotated data for highly accurate RNA structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022. doi: 10.48550/arXiv.2204.00300. URL https://arxiv.org/abs/2204.00300.

Weiqing Chen, Pengzhi Zhang, Tu N. Tran, Yiwei Xiao, Shengyu Li, Vrutant V. Shah, Hao Cheng, Kristopher W. Brannan, Keith Youker, Li Lai, Longhou Fang, Yu Yang, Nhat-Tu Le, Jun-ichi Abe, Shu-Hsia Chen, Qin Ma, Ken Chen, Qianqian Song, John P. Cooke, and Guangyu Wang. A visual–omics foundation model to bridge histopathology with spatial transcriptomics. *Nature Methods*, 22(7):1568–1582, 2025. doi: 10.1038/s41592-025-02707-1. URL https://doi.org/10.1038/s41592-025-02707-1.

Xingjian Chen, Jiecong Lin, Yuchen Wang, Weitong Zhang, Weidun Xie, Zetian Zheng, and Ka-Chun Wong. HE2Gene: Image-to-RNA translation via multi-task learning for spatial transcriptomics data. *Bioinformatics*, 40(6):btae343, 2024. doi: 10.1093/bioinformatics/btae343. URL https://doi.org/10.1093/bioinformatics/btae343.

Yiqun Chen and James Zou. Simple and effective embedding model for single-cell biology built from ChatGPT. *Nature Biomedical Engineering*, 9(4):483–493, 2025. doi: 10.1038/s41551-024-01284-6. URL https://doi.org/10.1038/s41551-024-01284-6.

Youngmin Chung, Ji Hun Ha, Kyeong Chan Im, and Joo Sang Lee. Accurate spatial gene expression prediction by integrating multi-resolution features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11591–11600, 2024. doi: 10.1109/CVPR52733.2024.01101. URL https://openaccess.thecvf.com/content/CVPR2024/html/Chung_Accurate_Spatial_Gene_Expression_Prediction_by_Integrating_Multi-Resolution_Features_CVPR_2024_paper.html.

Francesco Cisternino, Sara Ometto, Soumick Chatterjee, Edoardo Giacopuzzi, Adam P. Levine, and Craig A. Glastonbury. Self-supervised learning for characterising histomorphological diversity and spatial RNA expression prediction across 23 human tissue types. *Nature Communications*, 15(1):5906, 2024. doi: 10.1038/s41467-024-50317-w. URL https://doi.org/10.1038/s41467-024-50317-w.

Xiaohang Fu, Yue Cao, Beilei Bian, Chuhan Wang, Dinny Graham, Nirmala Pathmanathan, Ellis Patrick, Jinman Kim, and Jean Yee Hwa Yang. Spatial gene expression at single-cell resolution from histology using deep learning with GHIST. *Nature Methods*, 22(9):1900–1910, 2025. doi: 10.1038/s41592-025-02795-z. URL https://doi.org/10.1038/s41592-025-02795-z.

Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4(8):827–834, 2020. doi: 10.1038/s41551-020-0578-x. URL https://doi.org/10.1038/s41551-020-0578-x.

Susu Hu, Qinghe Zeng, Nithya Bhasker, Jakob Nikolas Kather, and Stefanie Speidel. HistoPrism: Unlocking functional pathway analysis from pan-cancer histology via gene expression prediction. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=6dTHxb9JuA.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Wengong Jin, and Zhitao Ying. Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 25550–25565. PMLR, 2025a. URL https://proceedings.mlr.press/v267/huang25t.html.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Rex Ying, and Wengong Jin. STPath: A generative foundation model for integrating spatial transcriptomics and whole-slide images. *npj Digital Medicine*, 8(1):659, 2025b. doi: 10.1038/s41746-025-02020-3. URL https://doi.org/10.1038/s41746-025-02020-3.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964. doi: 10.1214/aoms/1177703732. URL https://doi.org/10.1214/aoms/1177703732.

Guillaume Jaume, Paul Doucet, Andrew H. Song, Ming Y. Lu, Cristina Almagro-Pérez, Sophia J. Wagner, Anurag J. Vaidya, Richard J. Chen, Drew F. K. Williamson, Ahrong Kim, and Faisal Mahmood. HEST-1k: A dataset for spatial transcriptomics and histology image analysis. In *Advances in Neural Information Processing Systems*, volume 37, pp. 53798–53833, 2024. doi: 10.52202/079017-1704. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/60a899cc31f763be0bde781a75e04458-Abstract-Datasets_and_Benchmarks_Track.html.

Yuran Jia, Junliang Liu, Li Chen, Tianyi Zhao, and Yadong Wang. THItoGene: A deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):bbad464, 2024. doi: 10.1093/bib/bbad464. URL https://doi.org/10.1093/bib/bbad464.

Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W. King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, Mika Sarkin Jain, Jun Sung Park, Lauma Ramona, Elizabeth Tuck, Anna Arutyunyan, Roser Vento-Tormo, Moritz Gerstung, Louisa James, Oliver Stegle, and Omer Ali Bayraktar. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology*, 40:661–671, 2022. doi: 10.1038/s41587-021-01139-4. URL https://doi.org/10.1038/s41587-021-01139-4.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6):417–425, 2015. doi: 10.1016/j.cels.2015.12.004. URL https://doi.org/10.1016/j.cels.2015.12.004.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://doi.org/10.1126/science.ade2574.

Marija Milacic, Deidre Beavers, Patrick Conley, Chuqiao Gong, Marc Gillespie, Johannes Griss, Robin Haw, Bijay Jassal, Lisa Matthews, Bruce May, Robert Petryszak, Eliot Ragueneau, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Ralf Stephan, Krishna Tiwari, Thawfeek Varusai, Joel Weiser, Adam Wright, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The Reactome pathway knowledgebase 2024. *Nucleic Acids Research*, 52(D1): D672–D678, 2024. doi: 10.1093/nar/gkad1025. URL https://doi.org/10.1093/nar/gkad1025.

Wenwen Min, Zhiceng Shi, Jun Zhang, Jun Wan, and Changmiao Wang. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Briefings in Bioinformatics*, 25(6):bbae551, 2024. doi: 10.1093/bib/bbae551. URL https://doi.org/10.1093/bib/bbae551.

Kalin Nonchev, Sebastian Dawo, Karina Silina, Viktor H. Koelzer, and Gunnar Rätsch. DeepSpot-M: A multimodal foundation model for transcriptome-wide virtual spatial transcriptomics from histology. *medRxiv*, 2026. doi: 10.64898/2026.06.19.26356060. URL https://www.medrxiv.org/content/10.64898/2026.06.19.26356060v1.

Clare Pacini, Emma Duncan, Emanuel Gonçalves, James Gilbert, Shriram Bhosle, Stuart Horswell, Emre Karakoc, Howard Lightfoot, Ed Curry, Francesc Muyas, Monsif Bouaboula, Chandra Sekhar Pedamallu, Isidro Cortes-Ciriano, Fiona M. Behan, Lykourgos-Panagiotis Zalmas, Andrew Barthorpe, Hayley Francies, Steve Rowley, Jack Pollard, Pedro Beltrao, Leopold Parts, Francesco Iorio, and Mathew J. Garnett. A comprehensive clinically informed map of dependencies in cancer cells and framework for target prioritization. *Cancer Cell*, 42(2):301–316.e9, 2024. doi: 10.1016/j.ccell.2023.12.016. URL https://doi.org/10.1016/j.ccell.2023.12.016.

Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv*, 2021. doi: 10.1101/2021.11.28.470212. URL https://doi.org/10.1101/2021.11.28.470212.

Jiahe Qian, Yaoyu Fang, Ziqiao Weng, Xinkun Wang, Lee A. Cooper, and Bo Zhou. Learning from gene names, expression values and images: Contrastive masked text-image pretraining for spatial transcriptomics representation learning, 2025. URL https://arxiv.org/abs/2509.16892.

Md Mamunur Rahaman, Ewan K. A. Millar, and Erik Meijering. Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning. *Scientific Reports*, 13:13604, 2023. doi: 10.1038/s41598-023-40219-0. URL https://www.nature.com/articles/s41598-023-40219-0.

Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, Thomas Clozel, Matahi Moarii, Pierre Courtiol, and Gilles Wainrib. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature Communications*, 11(1):3877, 2020. doi: 10.1038/s41467-020-17678-4. URL https://doi.org/10.1038/s41467-020-17678-4.

Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul Igor Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg, and Jonas Frisén. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294): 78–82, 2016. doi: 10.1126/science.aaf2403. URL https://doi.org/10.1126/science.aaf2403.

Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L. Gable, Tao Fang, Nadezhda T. Doncheva, Sampo Pyysalo, Peer Bork, Lars J. Jensen, and Christian von Mering. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023. doi: 10.1093/nar/gkac1000. URL https://doi.org/10.1093/nar/gkac1000.

The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2026. *Nucleic Acids Research*, 54(D1):D1779–D1792, 2026. doi: 10.1093/nar/gkaf1292. URL https://doi.org/10.1093/nar/gkaf1292.

Chuhan Wang, Adam S. Chan, Xiaohang Fu, Shila Ghazanfar, Jinman Kim, Ellis Patrick, and Jean Y. H. Yang. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications*, 16(1):1544, 2025a. doi: 10.1038/s41467-025-56618-y. URL https://doi.org/10.1038/s41467-025-56618-y.

Hongyi Wang, Xiuju Du, Jing Liu, Shuyi Ouyang, Yen-Wei Chen, and Lanfen Lin. M2OST: Many-to-one regression for predicting spatial transcriptomics from digital pathology images. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(7):7709–7717, 2025b. doi: 10.1609/aaai.v39i7.32830. URL https://doi.org/10.1609/aaai.v39i7.32830.

P. Went, M. Vasei, L. Bubendorf, L. Terracciano, L. Tornillo, U. Riede, J. Kononen, R. Simon, G. Sauter, and P. A. Baeuerle. Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. *British Journal of Cancer*, 94:128–135, 2006. doi: 10.1038/sj.bjc.6602924.

Antonio C. Wolff, Mark R. Somerfield, Mitchell Dowsett, M. Elizabeth H. Hammond, David F. Hayes, Lisa M. McShane, Thomas J. Saphner, Patricia A. Spears, and Kimberly H. Allison. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology–College of American Pathologists guideline update. *Archives of Pathology & Laboratory Medicine*, 147(9):993–1000, 2023. doi: 10.5858/arpa.2023-0950-SA. URL https://doi.org/10.5858/arpa.2023-0950-SA.

Qitian Wu, Qiyu Gong, Luezhen Yuan, Zhiyi Li, Orr Ashenberg, Fei Chen, Ramnik Xavier, and Caroline Uhler. Multi-modal foundation model with whole-slide attention enables transferrable digital pathology at single-cell resolution. *bioRxiv*, 2026. doi: 10.64898/2026.07.31.741265. URL https://www.biorxiv.org/content/10.64898/2026.07.31.741265v1.

Xiao Xiao, Yan Kong, Ronghan Li, Zuoheng Wang, and Hui Lu. Transformer with convolution and graph-node co-embedding: An accurate and interpretable vision backbone for predicting gene expressions from local histopathological image. *Medical Image Analysis*, 91:103040, 2024. doi: 10.1016/j.media.2023.103040. URL https://doi.org/10.1016/j.media.2023.103040.

Ronald Xie, Kuan Pang, Sai W. Chung, Catia T. Perciani, Sonya A. MacParland, Bo Wang, and Gary D. Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 70626–70637, 2023. doi: 10.52202/075280-3095. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/df656d6ed77b565e8dcdfbf568aead0a-Abstract.html.

Ying Xiong, Linjing Liu, Yufei Cui, Shangyu Wu, Xue Liu, Antoni B. Chan, and Chun Jason Xue. GeneQuery: A general QA-based framework for spatial gene expression predictions from histology images, 2024. URL https://arxiv.org/abs/2411.18391.

Yan Yang, Md Zakir Hossain, Xuesong Li, Shafin Rahman, and Eric Stone. Spatial transcriptomics analysis of zero-shot gene expression prediction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15004 of *Lecture Notes in Computer Science*, pp. 492–502. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72083-3_46. URL https://papers.miccai.org/miccai-2024/718-Paper2573.html.

Yan Yang, Xuesong Li, Liyuan Pan, Guoxun Zhang, Liu Liu, and Eric Stone. AGP-Net: A universal network for gene expression prediction of spatial transcriptomics. *bioRxiv*, 2025. doi: 10.1101/2025.03.09.642276. URL https://doi.org/10.1101/2025.03.09.642276.

Wang Yin, Qin Peng, Fanyi Meng, You Wan, Weilong Zhang, and Yuan Zhou. HESpotEx: A dual-stream deep learning framework for spot-level gene expression prediction from histological images. *Nature Computational Science*, 6(8):842–855, 2026. doi: 10.1038/s43588-026-00992-0. URL https://doi.org/10.1038/s43588-026-00992-0.

Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, 2022. doi: 10.1093/bib/bbac297. URL https://doi.org/10.1093/bib/bbac297.

Chengming Zhang, Yiwen Yang, Yifeng Jiao, Qianqian Yang, Xin Guo, Jing Xu, Jiyang Li, Yuanyuan Zhou, Zhen Liu, Yushuai Wu, Yu Zhang, Chen Jiang, Jianyun Ma, Yuxuan Liu, Yao Zhang, Kai Xiao, Yidi Sun, Cirong Liu, Limei Han, Yuan Qi, Kazuyuki Aihara, Chengyu Li, Yuan Cheng, Luonan Chen, and Wu Wei. BrainBeacon: A cross-species foundation model for single-cell resolved brain spatial transcriptomics. *bioRxiv*, 2025. doi: 10.1101/2025.07.08.663729. URL https://doi.org/10.1101/2025.07.08.663729.

Daiwei Zhang, Amelia Schroeder, Hanying Yan, Haochen Yang, Jian Hu, Michelle Y. Y. Lee, Kyung S. Cho, Katalin Susztak, George X. Xu, Michael D. Feldman, Edward B. Lee, Emma E. Furth, Linghua Wang, and Mingyao Li. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, 42(9):1372–1377, 2024. doi: 10.1038/s41587-023-02019-9. URL https://doi.org/10.1038/s41587-023-02019-9.

Jeffrey Zhong, Lechuan Li, Ruth Dannenfelser, and Vicky Yao. Benchmarking gene embeddings from sequence, expression, network, and text models for functional prediction tasks. *bioRxiv*, 2025. doi: 10.1101/2025.01.29.635607. URL https://doi.org/10.1101/2025.01.29.635607.

Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion generative modeling for spatially resolved gene expression inference from histology images. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=FtjLUHyZAO.

Eric Zimmermann, Eugene Vorontsov, Julian Viret, Adam Casson, Michal Zelechowski, George Shaikovski, Neil Tenenholtz, James Hall, David Klimstra, Razik Yousfi, Thomas Fuchs, Nicolo Fusi, Siqi Liu, and Kristen Severson. Virchow2: Scaling self-supervised mixed magnification models in pathology, 2024. URL https://arxiv.org/abs/2408.00738.

# A Detailed model specification

This section gives the full mathematical specification underlying the model description. The notation uses spatial location $b$, cancer context $c$, requested gene $g$, latent width $D = 512$, spatial batch size $B$, and requested-gene chunk size $G$. Source-availability masks and expression-availability masks are distinct throughout.

## A.1 Problem formulation

Given a registered multiscale H&E region $I_b$, molecular priors $\mathcal{K}_g$, and cancer-context package $\mathcal{P}_c$, the model predicts one value for each gene–location pair:

$$\widehat{y}_{b,c,g} = f_\theta(I_b, \mathcal{K}_g, \mathcal{P}_c), \tag{6}$$

with parameters shared across genes and cancer contexts. A gene is queryable when it maps to the fixed identity registry and has at least one available molecular embedding; evaluation additionally requires a measured expression target. Exclusion from spatial-expression supervision therefore does not imply exclusion from external sequence, text, network, or pathway resources. Unmeasured expression values are masked rather than set to zero.

## A.2 Source-aware gene representations

Frozen ESM2 protein-sequence, RNA-FM RNA-sequence, and GenePT functional-text embeddings are projected separately to the common latent dimension:

$$e_{g,m} = \mathrm{LN}\big(P_m(\mathrm{embedding}_{g,m})\big)\,. \tag{7}$$

For source-availability indicator $A_{g,m} \in \{0, 1\}$, unavailable sources are excluded from masked attention. The fused representation is

$$\alpha_{g,m} = \frac{A_{g,m} \exp a_m(e_{g,m})}{\sum_{m'} A_{g,m'} \exp a_{m'}(e_{g,m'})}, \qquad q_g^0 = \mathrm{LN}\left(\sum_m \alpha_{g,m} e_{g,m} + e_{\text{gene-type},g}\right), \tag{8}$$

where $a_m$ is source-specific and the gene-type embedding is shared within protein_coding, other GENCODE biotypes, and unknown or unspecified categories. No per-gene identifier embedding is learned and missing source vectors are not synthesized.

**Cancer-independent relational refinement.** For normalized weights $w_{g,h}$ over a top-$k$ STRING neighborhood,

$$r_{\mathrm{STRING},g} = \sum_{h \in N_{\mathrm{STRING}}(g)} w_{g,h} W_{\mathrm{STRING}} q_h^0, \tag{9}$$

$$q_{\mathrm{shared},g} = \mathrm{LN}\big(q_g^0 + \gamma_{\mathrm{STRING}} r_{\mathrm{STRING},g}\big)\,. \tag{10}$$

The residual preserves the initial representation; genes without neighbors receive no graph message. The scalar $\gamma_{\mathrm{STRING}}$ is initialized to $0.05$.

## A.3 Cancer-context adaptation

The DepMap branch aggregates the cancer-context graph selected for $c$:

$$r_{\mathrm{DepMap},c,g} = \sum_{h \in N_{\mathrm{DepMap}}(c,g)} w_{c,g,h} W_{\mathrm{DepMap}} q_{\mathrm{shared},h}. \tag{11}$$

The cancer label selects the graph but is not embedded or passed directly to the decoder; a missing graph or gene gives a zero residual. The program branch has six source-separated banks: Cancer Card, Hallmark, Gene Ontology, Reactome, JUMP ORF, and JUMP CRISPR. Each bank has a separate projection, attention mask, null token, and gate. For source $m$,

$$r_{m,c,g} = \mathrm{CrossAttention}(q_{\mathrm{shared},g}, K_{m,c}, V_{m,c}), \tag{12}$$

$$\beta_{m,c,g} = v_{m,c}\ \mathrm{sigmoid}(G_m(q_{\mathrm{shared},g})), \tag{13}$$

where $v_{m,c} = 1$ only for a valid non-null token. Unavailable banks contribute no residual, and generic fallback banks are used only where explicitly defined. Program tokens encode external knowledge or perturbation-derived profiles, not measured activity in the target section. The cancer residual and image-reading query are

$$\Delta q_{\mathrm{cancer},c,g} = \gamma_{\mathrm{DepMap}} r_{\mathrm{DepMap},c,g} + \sum_m \eta_m \beta_{m,c,g} r_{m,c,g}, \tag{14}$$

$$q'_{c,g} = \mathrm{LN}(q_{\mathrm{shared},g} + \lambda_{\mathrm{ctx}} \Delta q_{\mathrm{cancer},c,g}). \tag{15}$$

The bounded shared scalar $\lambda_{\mathrm{ctx}}$ is initialized to $0.08$; $\gamma_{\mathrm{DepMap}}$ and $\eta_m$ are initialized to $0.05$. If neither a graph nor a program source is available, $\Delta q_{\mathrm{cancer},c,g} = 0$.

### A.4 Gene-guided multiscale histology reading

At each target spot, three registered contexts, R0, R1, and R2, are encoded by a shared frozen Virchow2 encoder (Zimmermann et al., 2024). Their anchor-neighborhood capacities are 7, 32, and 64 entries, respectively; these counts are not physical radii. Each valid anchor supplies 256 unpooled tokens of width 1,280, while invalid entries are masked. Scale adapters add learned scale embeddings and two-dimensional positions, shared spatial self-attention contextualizes the tokens, and a learned 16-query resampler produces

$$X_{\mathrm{image}} \in \mathbb{R}^{B\times 3\times 16\times 512}. \tag{16}$$

The resampler retains attention weights over original patch coordinates for visualization. For scale $s$, image keys and values are computed solely from H&E-derived latents:

$$K_{\mathrm{image},s} = W_K X_s, \qquad V_{\mathrm{image},s} = W_V X_s. \tag{17}$$

Molecular and program tokens are not inserted into these key–value sequences. Image geometry, cache construction, and invalid-anchor handling are specified in Sections F and G.

**Gene-conditioned image readout.** The cancer-adapted query cross-attends to each scale:

$$z_{b,c,g,s} = \mathrm{CrossAttention}\big(Q = q'_{c,g}, K = K_{\mathrm{image},b,s}, V = V_{\mathrm{image},b,s}\big). \tag{18}$$

The readout has no additive query residual: its values originate in the image branch and the gene query determines their combination. The gene–location router is

$$r_{c,g} = \tanh\big(W_{\mathrm{route}} q'_{c,g}\big), \tag{19}$$

$$\ell_{b,c,g,s} = F_{\mathrm{scale}}([z_{b,c,g,s},\ z_{b,c,g,s} \odot r_{c,g}]), \tag{20}$$

$$w_{b,c,g,s} = \mathrm{softmax}_s(\ell_{b,c,g,s}), \tag{21}$$

$$z_{\mathrm{image},b,c,g} = \sum_s w_{b,c,g,s} z_{b,c,g,s}. \tag{22}$$

The weights depend jointly on the adapted gene query and scale-specific image readouts, without manually assigning a cancer-level scale preference.

### A.5 Shared cancer-conditioned expression decoder

Shared conditioner networks transform the shared gene representation and cancer residual:

$$u_{\mathrm{shared},g} = F_{\mathrm{shared}}(q_{\mathrm{shared},g}), \tag{23}$$

$$u_{\mathrm{cancer},c,g} = F_{\mathrm{cancer}}(\Delta q_{\mathrm{cancer},c,g}). \tag{24}$$

Cancer modulation is bounded by

$$\lambda_m = \lambda_{\max} \operatorname{sigmoid}(\rho_m), \tag{25}$$

with effective initialization near 0.05–0.1. Feature-wise modulation retains the unmodulated image path:

$$m_{c,g} = \tanh(u_{\text{shared},g} + \lambda_m u_{\text{cancer},c,g}), \tag{26}$$

$$\widetilde{z}_{b,c,g} = z_{\text{image},b,c,g} \odot m_{c,g}, \tag{27}$$

$$h_{b,c,g} = \operatorname{concat}(z_{\text{image},b,c,g}, \widetilde{z}_{b,c,g}). \tag{28}$$

The final prediction is

$$\widehat{y}_{b,c,g} = \text{SharedDecoder}(h_{b,c,g}). \tag{29}$$

The shared decoder receives the image readout and modulated counterpart, not the cancer label or a separate raw molecular input. Its parameter count is independent of the requested gene panel.

### A.6 Training

Expression targets are library-size normalized and log transformed:

$$x'_{b,g} = \log\left(1 + 10^4 \frac{x_{b,g}}{\sum_{g'} x_{b,g'}}\right). \tag{30}$$

We use $y_{b,g} = x'_{b,g}$ without gene-specific target standardization. Stage 1 minimizes masked Huber loss with $\delta = 1.0$ (Huber, 1964):

$$\mathcal{L}_{\text{stage1}} = \frac{\sum_{b,g} M_{b,g} \operatorname{Huber}(\widehat{y}_{b,g}, y_{b,g})}{\sum_{b,g} M_{b,g}}. \tag{31}$$

Stage 2 combines masked half-MSE with within-section gene-wise PCC:

$$\mathcal{L}_{\text{stage2}} = \frac{\sum_{b,g} M_{b,g} \frac{1}{2}(\widehat{y}_{b,g} - y_{b,g})^2}{\sum_{b,g} M_{b,g}} + \lambda_{\text{PCC}} \frac{\sum_{(s,g)\in\mathcal{V}} [1 - \text{PCC}_{s,g}]}{|\mathcal{V}|}, \qquad \lambda_{\text{PCC}} = 0.1. \tag{32}$$

The set $\mathcal{V}$ contains gene–section groups with at least eight valid spots and target variance above $10^{-8}$; at most 128 spots contribute to a group. Correlations are computed within sections, never by pooling unrelated sections or cancer contexts. Genes without an eligible group contribute only to masked value loss, and no auxiliary loss is used.

Spatial batches are cancer-balanced and genes are processed in chunks of at most 256. The gene schedule allocates its draw budget as

$$\pi_{\text{uniform}} = 0.8, \qquad \pi_{\text{priority},c} = 0.2. \tag{33}$$

At the start of each epoch, a cancer-specific schedule is constructed from 35,135 eligible genes after excluding $\mathcal{U}_{200}$. Its 35,135 draw positions comprise 7,027 cancer-priority draws and 28,108 uniform draws from the remaining eligible genes. The two pools are shuffled deterministically using the training seed, epoch, and cancer. A cumulative quota assigns approximately 20% of each chunk to the priority pool; gene IDs are unique within a chunk but can recur across chunks. The 138 resulting chunks are cycled over that cancer's spatial minibatches. Thus, $0.8/0.2$ specifies draw allocation, and constructing the schedule once per epoch does not imply that every gene is trained exactly once.

Virchow2, the three molecular encoders, and precomputed program embeddings remain frozen; source projectors, graph adapters, program attention, image adapters, spatial attention, resampling, routing, conditioner networks, the shared decoder, and residual/modulation coefficients are trained jointly. With $S = 3$ scales and $L = 16$ latents per scale, gene–image cross-attention has an approximate cost $\mathcal{O}(BGSLD)$; chunking bounds peak memory by $G_{\text{chunk}}$.

## B Detailed gene-source preprocessing

The gene branch uses three frozen embedding sources with separate trainable projections: ESM2 (5,120 dimensions), RNA-FM (640), and GenePT (1,536). Table A1 lists the prepared matrices and

Table A1: **Prepared molecular-prior matrices.** Source vectors are mapped to the 78,733-gene registry. Stored matrix size and available gene coverage are distinct; availability masks determine which rows can contribute to a gene query.

| Source | Stored shape | Available rows |
|---|---|---|
| ESM2 (BrainBeacon-prepared) | $21{,}096 \times 5{,}120$ | 18,547 |
| RNA-FM | $78{,}733 \times 640$ | 58,626 |
| GenePT | $78{,}733 \times 1{,}536$ | 33,015 |

their available rows. The ESM2 matrix was prepared using the BrainBeacon resource pipeline (Zhang et al., 2025). Source preparation and coverage are detailed below.

ESM2 vectors are retained for genes annotated `protein_coding` with unique versionless Ensembl matches in GENCODE v50. Reconciliation excludes 135 rows with other GENCODE biotypes and 2,414 identifiers outside the registry. RNA-FM uses a frozen 12-layer, 640-dimensional encoder with deterministic canonical-transcript selection from GENCODE v50. Sequences are divided into windows of at most 1,022 nucleotides with 128-nucleotide overlap; overlapping positions are coverage-corrected before gene-level averaging, and special tokens are excluded. The prepared RNA-FM availability mask covers 58,626 entries with other GENCODE biotypes, including annotated pseudogenes, TEC entries, and IG/TR genes. This annotation category is not synonymous with biological noncoding genes, and sequence availability does not imply that every entry has a characterized biological function.

GenePT vectors are taken from Zenodo record 10833191 and matched to case-insensitive canonical GENCODE gene names. Ambiguous matches and unmatched names remain unavailable; aliases are not used to resolve them automatically. Missing vectors are not generated from target expression.

## C IDENTIFIER MAPPING AND SOURCE MASKS

The master registry contains 78,733 human genes from GENCODE v50 on GRCh38.p14, ordered by versionless Ensembl gene ID. GENCODE `gene_name` supplies the display name; HGNC identifiers and aliases are retained as cross-references, not automatic replacements. Three loci have ambiguous HGNC mappings in the source records. The order of requested genes is preserved independently of registry order.

The registry, source-available genes, expression-supervised genes, and evaluation panels define different sets. Gene-source availability is recorded separately for each source, whereas the spot–gene expression mask indicates whether a measurement exists. Unmeasured expression is not treated as a zero count.

## D STRING AND DEPMAP GRAPH CONSTRUCTION

The prepared graph uses human STRING v12 (taxon 9606), mapped to the gene registry and stored as normalized top-32 neighbor lists. It includes 19,131 mapped genes and 611,384 directed neighbor entries. The stage-1 configuration also specifies 32 neighbors.

DepMap graphs use the Public 26Q1 `CRISPRGeneEffect.csv` matrix, with cell-line metadata from `Model.csv` and Entrez-to-Ensembl mappings from `Gene.csv`. Only unique gene mappings into the registry are retained. Contexts use `OncotreePrimaryDisease`: Colorectal Adenocarcinoma (63 models), Renal Cell Carcinoma (29), Adult-Type Diffuse Glioma (69), and Invasive Breast Carcinoma (50). Prostate uses `OncotreeLineage=Prostate` (10 models). Counts refer to models represented in the CRISPR-effect matrix. A context requires at least 20 such models, so prostate has a null graph.

Within a context with $n$ models, a gene must have at least $\max(20, \lceil 0.8n \rceil)$ finite measurements. Profiles are centered by their observed means, missing entries are set to zero after centering, and nonconstant profiles are normalized to unit $\ell_2$ norm. The edge score is the profile dot product multiplied by the smaller observation fraction of the two genes. Self-edges and scores below $0.20$

are removed; at most 32 highest-scoring neighbors are retained per gene. During message passing, retained weights are divided by their sum for each target gene. The resulting padded neighbor array has shape $[5, 78{,}733, 32]$.

Both graph adapters use sparse neighborhoods. STRING refines the shared gene representation, while DepMap contributes to the cancer residual. Missing graph information gives a zero residual. The residual coefficients $\gamma_{\mathrm{STRING}}$ and $\gamma_{\mathrm{DepMap}}$ are unconstrained trainable scalars initialized to $0.05$.

## E Cancer Card and program-bank construction

Cancer knowledge is supplied as a cancer-level package shared across sections of the same context, rather than inferred from each test image. The Cancer Card collection contains 13 literature-derived programs per cancer, compiled from cancer literature published over the past 10 years. These programs describe key genes, pathways, and cancer-subtype context. Card embeddings condition gene queries; the separately derived gene-priority weights control the epoch-level training schedule, whereas the cancer-related panel is a reporting set. Card genes may overlap the cancer-related panel, and neither source directly supplies a target expression value.

Table A2: **Source-separated molecular program banks.** Counts include a null token. Cancer Cards report active tokens per context; the other rows report source inventories before cancer-dependent selection.

| Source | Rows | Native dimension | Content |
|---|---|---|---|
| Cancer Card | $5 \times 14$ | 768 | Thirteen literature-derived programs plus null per cancer |
| GO | 17,806 | 768 | Term names and definitions |
| Reactome | 2,312 | 768 | Human pathway names |
| Hallmark | 51 | 768 | MSigDB 2026.1.Hs, 50 sets plus null |
| JUMP ORF | 12,052 | 722 | Aggregated gene-perturbation profiles |
| JUMP CRISPR | 7,804 | 259 | Aggregated gene-perturbation profiles |

Text descriptions are encoded using frozen BiomedBERT/PubMedBERT representations, followed by source-specific trainable projections. GO contributes 17,805 non-null terms and Reactome 2,311 human pathways. Missing pathway descriptions are not synthesized. The Hallmark source uses MSigDB 2026.1.Hs. JUMP ORF and CRISPR retain separate perturbation-profile spaces, gene identifiers, replicate metadata, and quality-control information; the program banks do not contain raw microscopy images.

For each cancer, GO and Reactome candidates are ranked by overlap between their gene memberships and Cancer Card genes; JUMP profiles are ranked by the corresponding perturbation-gene overlap. These rankings use no target expression. The preselected lists contain 256 non-null candidates plus a null token per source. After duplicate removal, the model reads at most 64 tokens per source and cancer, including the null token. Thus GO, Reactome, JUMP ORF, and JUMP CRISPR each supply 64 tokens. Hallmark retains all 50 programs plus null, and Cancer Card supplies 13 programs plus null; its two additional storage slots are masked padding.

Each bank has a validity mask, null token, projection, and source-specific attention. Source-native vectors are not concatenated into a single bank. Program residual coefficients $\eta_m$ are unconstrained trainable scalars initialized to $0.05$.

## F PHYSICAL IMAGE SCALES AND PATCH EXTRACTION

Each image anchor supplies a $224 \times 224$-pixel crop at $0.5\,\mu\mathrm{m/pixel}$, corresponding to a $112 \times 112\,\mu\mathrm{m}$ field. Anchors lie on a hexagonal grid with $100\,\mu\mathrm{m}$ spacing. For each Visium coordinate, Euclidean distance assigns anchors to the annular bands $\mathrm{R0} = [0, 125)$, $\mathrm{R1} = [125, 500)$, and $\mathrm{R2} = [500, 2000)\,\mu\mathrm{m}$. Within each band, the nearest 7, 32, and 64 anchors are retained, respectively; missing slots are padded and masked. These capacities bound the number of image fields, while the annular limits define their spatial coverage. Each valid anchor yields a $16 \times 16$ grid of unpooled Virchow2 patch tokens (256 tokens of width 1,280). The learned resampler returns 16 latents per scale.

Invalid anchors are zeroed and masked. When an entire neighborhood is invalid, a dummy key prevents undefined attention weights and the resulting latents are set to zero. This numerical safeguard is distinct from the image-boundary policy used during patch extraction. The paired Visium prediction route does not require a dense whole-slide grid.

## G VIRCHOW2 TOKEN EXTRACTION AND CACHING

Frozen Virchow2 features are extracted once for each image field and reused during training and inference. The cache associates unpooled patch tokens with token coordinates, scale, section, and spot identity. This separates image encoding from optimization of the gene-conditioned model.

Cache checks verify row order, spot identity, scale count, tensor shape, and finite values before use.

## H ARCHITECTURE AND TENSOR DIMENSIONS

Table A3 summarizes the primary tensors for a batch of $B$ locations and a gene chunk of $G$ genes. Program-bank length varies by source and selected cancer context.

Table A3: **Principal tensor dimensions.** $B$ denotes spatial batch size, $G$ the requested gene chunk, and $D = 512$ the latent width.

| Tensor | Shape |
|---|---|
| Source-projected gene vectors $e_{g,m}$ | $[G, D]$ per source |
| Shared gene representation $q_{\mathrm{shared}}$ | $[B, G, D]$ after cancer broadcasting |
| Cancer residual $\Delta q_{\mathrm{cancer}}$ | $[B, G, D]$ |
| Image-reading representation $q'$ | $[B, G, D]$ |
| Resampled image tokens | $[B, 3, 16, D]$ |
| Per-scale image readout $z_{b,c,g,s}$ | $[B, G, 3, D]$ |
| Scale weights $w_{b,c,g,s}$ | $[B, G, 3]$ |
| Routed image readout $z_{\mathrm{image}}$ | $[B, G, D]$ |
| Decoder input $h$ | $[B, G, 2D]$ |
| Prediction $\widehat{y}$ | $[B, G]$ |

The conditioner networks use input and output LayerNorm around a residual two-layer MLP:

$$v = \mathrm{LN}_{\mathrm{in}}(x), \qquad F(x) = \mathrm{LN}_{\mathrm{out}}[v + W_2\,\mathrm{Dropout}_{0.1}(\mathrm{GELU}(W_1 v + b_1)) + b_2]\,, \tag{34}$$

where $W_1$ maps $512 \to 1024$ and $W_2$ maps $1024 \to 512$. LayerNorm is inside $F$, not applied a second time at its call site. These dimensions specify the shared and cancer-context modulation networks used by the decoder. The shared decoder is Linear$(1024, 512)$–GELU–Dropout–Linear$(512, 256)$–GELU–Linear$(256, 1)$.

The cancer conditioner contains bias terms, so $F_{\mathrm{cancer}}(0)$ need not be zero. Thus, absence of cancer priors sets $\Delta q_{\mathrm{cancer},c,g} = 0$ but does not necessarily remove every cancer-conditioner contribution to decoder modulation. The context strength parameters are bounded, unlike the graph and program residual coefficients.

Source weights, program gates, graph residual norms, image and resampler attention, scale weights, and modulation strengths can be recorded for analysis. These quantities describe model utilization and routing, not causal biological importance.

# I TRAINING ALGORITHM

**Algorithm 1** One GENEROUTE training step within the epoch-level sampling schedule

**Require:** cancer-balanced spatial batch $\mathcal{B}$; gene chunk $\mathcal{G}$; source masks $A$; expression mask $M$
*Epoch setup:* build one 80% uniform / 20% priority gene schedule per cancer; cycle its chunks over spatial minibatches
Load frozen ESM2, RNA-FM, and GenePT vectors for $\mathcal{G}$
Project and availability-fuse the gene sources to obtain $q^0$
Apply the sparse STRING residual to obtain $q_{\text{shared}}$
Select the batch cancer package and apply the sparse DepMap residual
Read each source-separated program bank and construct $\Delta q_{\text{cancer}}$
Form the cancer-adapted gene representation $q'$ using Eq. (15)
Load or extract frozen Virchow2 tokens for $\mathrm{R0}$, $\mathrm{R1}$, and $\mathrm{R2}$
Apply scale adapters, positions, shared spatial self-attention, and token resampling
Cross-attend each cancer-adapted gene representation to each scale's pure image keys and values
Compute gene–spot scale weights and obtain $z_{\text{image}}$
Compute shared and cancer modulation and form $h$
Predict one scalar per gene with the shared decoder
Compute the stage-specific masked expression objective; in stage 2 add the same-section gene-PCC term for groups of at most 128 spots
Update every trainable module; keep the frozen encoders and embeddings fixed

Distributed workers synchronize cancer-balanced sampling and requested gene chunks. The reported run uses eight ranks, 32 spots per rank, and bfloat16 mixed precision. Checkpoints are recorded at epoch boundaries.

# J SELECTION OF SUPERVISION-UNSEEN GENES

The fixed held-out set $\mathcal{U}_{200}$ contains 200 versionless Ensembl identifiers. All GeneRoute expression training excludes these genes from the start of stage 1, including the checkpoint used for continuation and both training objectives. Validation scores, early stopping, and hyperparameter selection also exclude them. External molecular priors remain available. Thus, held-out prediction tests transfer beyond expression supervision; the reference-access differences in Table A7 still apply.

**Selection and reproducibility.** Selection is stratified by GENCODE annotation biotype and molecular-prior availability within a restricted candidate pool with shared assay coverage. This pool contains 13,144 registry-mapped genes after excluding a pre-existing 2,000-gene panel from the 15,144 shared identifiers. The held-out panel includes all 62 RNA-FM-available candidates with other GENCODE biotypes in this pool. From ESM2-available `protein_coding` candidates, 69 genes are selected with GenePT and 69 without GenePT. Within each `protein_coding` pool, candidates are ranked by the ascending SHA-256 digest of `seed|versionless_Ensembl_ID`, and the first 69 are selected. The fixed seed is `generoute_st_v1_permanent_heldout_20260901`. Of the 62 genes with other GENCODE biotypes, 20 have GenePT and 42 do not (Table A4). This design emphasizes complementary prior availability rather than matching population gene-type proportions. The complete list and stratum assignments appear in Table S5 and the accompanying `u200_genes.csv`.

**RNA-FM supervision outside the held-out panel.** The selection pool is narrower than the expression-training vocabulary. After held-out exclusion, the latter contains 35,135 eligible genes, including 16,199 RNA-FM-available genes with other GENCODE biotypes. When sampled with measured targets, these genes contribute to the expression loss. Their frozen RNA-FM embeddings pass through the trainable source projection and availability-masked fusion into the shared prediction path. The projection therefore has a direct supervised loss path through non-held-out genes; its training does not depend solely on indirect graph connections to the 62 held-out genes.

**Supervised-panel construction and overlap.** The four panels are frozen separately for each cancer using training data, after excluding $\mathcal{U}_{200}$. High-variance genes are ranked by variance of library-

Table A4: **Composition of the permanent held-out panel.** P denotes genes annotated `protein_coding` with ESM2; N denotes genes with other GENCODE biotypes and RNA-FM. +/− indicates GenePT availability. The N groups exhaust the restricted held-out candidate pool, not the RNA-FM-available training vocabulary.

| Gene class | Sequence source | GenePT available | GenePT unavailable |
|---|---|---|---|
| `protein_coding` | ESM2 | 69 (P+) | 69 (P−) |
| Other GENCODE biotypes | RNA-FM | 20 (N+) | 42 (N−) |
| Total | | 89 | 111 |

normalized $\log(1 + x)$ expression; low-variance genes are ranked in the opposite direction after requiring detection in at least 5% of spots in at least 80% of training sections. HVGs use the Seurat-v3 selection procedure. The cancer-related panel ranks available marker candidates by cancer-specific priority, with a nominal quota of 100 and 60–68 selected genes per cancer. The nominal quotas are 1,000 high-variance, 500 low-variance, and 1,000 HVG genes. The selected source identifiers are mapped to queryable identifiers before scoring; the complete membership lists retain both axes. Across cancers, the selected union contains 6,100 identifiers and the queryable union contains 6,088.

For every cancer $c$ and panel $p$, $\mathcal{P}_{c,p} \cap \mathcal{U}_{200} = \varnothing$. This was checked directly on both selected and queryable identifier sets. The supervised panels can overlap each other: for example, high-variance and HVG panels share 205–297 selected genes per cancer (Table A5). These are overlapping views of supervised prediction, not mutually exclusive test categories. The membership and intersection records are supplied as `supervised_panel_membership.csv` and `gene_panel_audit.json`.

Table A5: **Pairwise overlaps among supervised panels.** Counts use the frozen selected source identifiers. HV, LV, and CR denote high-variance, low-variance, and cancer-related panels. Every supervised panel has zero overlap with U200 on both the selected and queryable axes.

| Cancer | HV∩LV | HV∩HVG | HV∩CR | LV∩HVG | LV∩CR | HVG∩CR |
|---|---|---|---|---|---|---|
| Brain/glioma | 0 | 208 | 15 | 4 | 0 | 32 |
| Breast | 0 | 205 | 23 | 1 | 2 | 34 |
| Colorectal | 0 | 297 | 24 | 2 | 1 | 45 |
| Kidney RCC | 0 | 250 | 27 | 2 | 1 | 37 |
| Prostate | 0 | 248 | 17 | 3 | 0 | 26 |

## K HYPERPARAMETERS

The pooled model uses 60 epochs of stage-1 Huber training followed by 14 epochs of the stage-2 half-MSE/PCC objective. Each epoch has 1,252 optimizer updates, giving 75,120 stage-1 updates and 17,528 further updates, for 92,648 in total. The objective switches after stage 1 while retaining optimizer and random-number states. Training uses eight ranks, 32 spots per rank, bfloat16 mixed precision, and seed 20260903. Table A6 summarizes the settings.

AdamW uses a base learning rate of $10^{-4}$ for source, image, and decoder modules and $5 \times 10^{-5}$ for graph and program modules, with zero weight decay. A shared multiplier warms up linearly over the first 2,504 updates and then follows cosine decay to $0.01$ at update 25,040. Over the next 1,252 updates it rises linearly from $0.01$ to $0.30$, then follows a cosine decay toward $0.01$ at the 125,200-update schedule endpoint. The reported checkpoint at 92,648 updates lies on this final cosine segment of the configured 125,200-update horizon. The schedule continues across the change of objective without resetting the optimizer.

Table A6: **Architecture and training settings for the pooled model.**

| Hyperparameter | Value |
|---|---|
| Latent dimension $D$ | 512 |
| Attention heads | 8 |
| Spatial self-attention layers | 2 |
| Program cross-attention layers | 1 |
| Image cross-attention layers | 1 |
| Image tokens per scale | 16 |
| Scales | R0, R1, R2 |
| Gene chunk size | 256 |
| $\lambda_{\mathrm{ctx}}$ initialization | 0.08 |
| $\lambda_m$ initialization | 0.08 |
| $\lambda_{\mathrm{ctx}}$ bound and parameterization | $\lambda_{\mathrm{ctx,max}}\ \mathrm{sigmoid}(\rho_{\mathrm{ctx}})$, maximum 1.0 |
| Gene sampling mixture | 80% uniform / 20% cancer-priority draw budget; schedule rebuilt each epoch |
| Gradient clipping | Global norm 1.0 |
| Virchow2 | frozen |
| External gene/program embeddings | frozen |
| Optimizer and base learning rates | AdamW; $10^{-4}$, with $5 \times 10^{-5}$ for graph/program modules |
| Weight decay | 0 |
| Batch size | 32 spots per rank; 8 ranks |
| Mixed precision | bfloat16 |
| Selected checkpoint | GeneRoute pooled checkpoint at 92,648 updates; continuation from the earlier stage-1 checkpoint |
| $\lambda_{\max}$ | 1.0 (configuration bound) |
| Stage-2 expression/PCC objective | half-MSE + 0.1 same-section gene-wise PCC; minimum 8 valid spots |

## L BASELINE IMPLEMENTATION DETAILS

All methods use the same benchmark training, validation, and test partitions and the same scored gene panels. They retain their respective architectures and pretrained resources, while reference and target conventions remain method-specific. All benchmark methods use a maximum of 100 training epochs and validation-based early stopping. Component ablations use a fixed budget of 20 epochs of stage 1.

GeneQuery uses a native target defined separately for each gene and section. Let $x_{s,i,g}$ denote the raw measured count for spot $i$, gene $g$, and section $s$, and let $\mathcal{I}_{s,g}$ be the spots with a valid measurement. No library-size normalization is applied. Its native target is

$$\begin{aligned} u_{s,i,g} &= \log(1 + x_{s,i,g}), \\ \widetilde{x}^{\mathrm{native}}_{s,i,g} &= \frac{u_{s,i,g} - \min_{j \in \mathcal{I}_{s,g}} u_{s,j,g}}{\max_{j \in \mathcal{I}_{s,g}} u_{s,j,g} - \min_{j \in \mathcal{I}_{s,g}} u_{s,j,g} + 10^{-8}}, \quad i \in \mathcal{I}_{s,g}. \end{aligned} \tag{35}$$

Missing measurements are excluded from both the extrema and the loss. We report GeneQuery gene-wise correlations against this native target and omit common-target spot-wise metrics and RMSE. The affine step preserves within-gene correlations with the corresponding $\log(1 + x)$ values; it is not equivalent to the common library-normalized target. The § label identifies these native-target reference results, which are excluded from common-target best/second-best ranking. The measured-expression reference and held-out information access of TissueFormer and SGN, and the H&E-plus-prior access of GENEROUTE, are summarized in Supplementary Table A7. DeepSpot-M uses externally prepared priors.

Table A7: **Inference information access.** TissueFormer uses measured expression from training reference spots; SGN uses a training reference only for held-out-gene calibration. None of these methods accesses ground-truth expression from the scored test sections.

| Method | Inference inputs and reference use | Held-out expression access |
|---|---|---|
| TissueFormer† | In-context retrieval without an expression decoder. Predictions are softmax-weighted measured expression from the top 64 similar training-reference spots. | The training reference includes measured expression for all 200 held-out genes. |
| SGN | Supervised panels: direct common-target training. $\mathcal{U}_{200}$: gene-wise $\log_{10}$-count calibration followed by library normalization (Eq. (36)); no reference-spot retrieval. | Calibration uses held-out gene ranges from 17 training sections (41,173 spots) and their median library size of 7,783. |
| GENEROUTE | H&E and external gene priors, without measured-expression references. | $\mathcal{U}_{200}$ is excluded from both expression-training stages, validation scores, early stopping, and hyperparameter selection. |

**SGN held-out calibration.** For $\mathcal{U}_{200}$ only, let $p_{s,i,g} \in [0,1]$ be the native SGN output. The per-gene reference extrema $a_g$ and $b_g$ are the minimum and maximum of $\log_{10}(1 + x_{s,i,g})$ over valid measurements in the same 17 training sections (41,173 spots); the median training-reference library size is 7,783. The calibration is

$$\begin{aligned} z_{s,i,g} &= p_{s,i,g}(b_g - a_g) + a_g, \\ r_{s,i,g} &= \max(\exp(\ln(10)\, z_{s,i,g}) - 1, 0)\,, \\ y^{\mathrm{SGN}}_{s,i,g} &= \log\left(1 + \frac{10^4 r_{s,i,g}}{7783}\right). \end{aligned} \tag{36}$$

The extrema and library size use training references only; test expression is never used. Missing reference measurements are excluded from the extrema.

### L.1 METRIC DEFINITIONS AND AGGREGATION

We report spot-wise PCC, spot-wise Spearman's $\rho$, gene-wise PCC, gene-wise Spearman's $\rho$, and root-mean-squared error (RMSE). Gene-wise metrics compare a gene's predicted and observed expression across valid spots within a section, then average defined correlations over genes. Spot-wise metrics compare expression profiles across valid genes at each spot, then average defined correlations over spots. The Spearman metrics use the corresponding ranks. Section scores are averaged equally within each cancer, and cancer means are averaged equally for the five-cancer macro results. RMSE is averaged from section-level errors. A macro result is unavailable if a cancer lacks a defined score.

For every section-level correlation, a value is undefined when fewer than two valid prediction–truth pairs remain or when $\sqrt{\mathrm{var}(\widehat{y})\,\mathrm{var}(y)} \le 10^{-12}$. The prediction and truth variance check is symmetric for gene-wise correlations across spots and spot-wise correlations across genes. Undefined values are represented as NaN, excluded from the corresponding section mean, and never replaced by zero; constant-gene and constant-spot counts are recorded. The same requested panel can therefore yield different defined-correlation subsets after method-specific constant masking. Higher correlations and lower RMSE indicate better performance. Macro-table standard deviations describe variation across the five cancer means; per-cancer standard deviations describe variation across test sections. They are not confidence intervals or variation across training seeds. GeneQuery's gene-wise correlations are shown on its native target, while its spot-wise correlations and RMSE are omitted from the common-target comparison. GeneQuery is evaluated against $\log(1 + x)$-derived native targets rather than the library-normalized targets used by the other methods (Eq. (35)). Its native-target values are excluded from common-target best/second-best ranking.

Table A8: **Cancer-related prediction by cancer context.** Spot-wise and gene-wise PCC for the same panel as Table 1. Values are means with standard deviations across test sections within each cancer.

| | $PCC_{spot}$ | | | | | $PCC_{gene}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | Brain | Breast | CRC | RCC | Pros. | Brain | Breast | CRC | RCC | Pros. |
| HE2Gene | 0.3132 (±0.128) | 0.6464 (±0.057) | 0.6358 (±0.071) | 0.5847 (±0.096) | 0.4869 (±0.094) | 0.0235 (±0.039) | 0.0277 (±0.047) | 0.0881 (±0.040) | 0.0293 (±0.062) | 0.0976 (±0.013) |
| ST-Net | 0.2962 (±0.036) | 0.6415 (±0.127) | 0.6194 (±0.089) | 0.6820 (±0.018) | 0.2864 (±0.063) | 0.0029 (±0.012) | -0.0011 (±0.049) | 0.0650 (±0.048) | 0.0062 (±0.042) | 0.0396 (±0.044) |
| HisToGene | 0.4932 (±0.128) | 0.6317 (±0.024) | 0.6644 (±0.054) | 0.7181 (±0.015) | 0.4749 (±0.027) | 0.0092 (±0.031) | 0.0521 (±0.000) | 0.0524 (±0.025) | 0.0029 (±0.015) | 0.0625 (±0.023) |
| Hist2ST | 0.4222 (±0.101) | 0.6514 (±0.179) | 0.5877 (±0.071) | 0.7072 (±0.051) | 0.4116 (±0.119) | -0.0003 (±0.001) | -0.0051 (±0.008) | 0.0034 (±0.007) | 0.0003 (±0.004) | 0.0037 (±0.003) |
| THItoGene | 0.4029 (±0.084) | 0.6544 (±0.137) | 0.5884 (±0.079) | 0.6794 (±0.031) | 0.3997 (±0.161) | 0.0014 (±0.002) | -0.0005 (±0.002) | -0.0029 (±0.012) | -0.0016 (±0.001) | 0.0048 (±0.005) |
| BrSTNet | 0.3303 (±0.060) | 0.6378 (±0.151) | 0.6039 (±0.086) | 0.6806 (±0.020) | 0.3467 (±0.113) | 0.0054 (±0.015) | 0.0205 (±0.063) | 0.0498 (±0.033) | -0.0158 (±0.014) | 0.0048 (±0.011) |
| DeepSpot-M | 0.3653 (±0.063) | 0.6748 (±0.145) | 0.5846 (±0.085) | 0.6806 (±0.033) | 0.3436 (±0.113) | 0.0121 (±0.044) | 0.0515 (±0.007) | 0.0333 (±0.051) | 0.0524 (±0.013) | 0.0064 (±0.033) |
| STFlow | 0.5629 (±0.331) | 0.7614 (±0.017) | 0.7085 (±0.065) | 0.7556 (±0.041) | 0.7481 (±0.113) | **0.0623** (±0.057) | 0.1169 (±0.061) | 0.1377 (±0.066) | 0.0806 (±0.058) | 0.1020 (±0.052) |
| TCGN | 0.3024 (±0.068) | 0.6557 (±0.133) | 0.4797 (±0.102) | 0.7076 (±0.034) | 0.3408 (±0.103) | 0.0129 (±0.028) | 0.0839 (±0.007) | 0.0415 (±0.023) | 0.0091 (±0.019) | 0.0593 (±0.054) |
| GeneQuery§ | – | – | – | – | – | -0.0008 (±0.021) | 0.0130 (±0.045) | 0.0345 (±0.071) | 0.0775 (±0.018) | 0.0353 (±0.053) |
| SGN | 0.3759 (±0.085) | 0.4734 (±0.291) | 0.3721 (±0.118) | 0.6509 (±0.067) | 0.1683 (±0.129) | 0.0229 (±0.047) | 0.0523 (±0.002) | -0.0149 (±0.038) | 0.0111 (±0.008) | 0.0025 (±0.020) |
| TissueFormer† | 0.3197 (±0.130) | 0.5687 (±0.215) | 0.5612 (±0.130) | 0.7003 (±0.042) | 0.7158 (±0.108) | 0.0083 (±0.003) | 0.0706 (±0.021) | 0.0945 (±0.030) | 0.0316 (±0.021) | 0.0718 (±0.027) |
| GeneRoute | **0.5878** (±0.276) | **0.7833** (±0.054) | **0.7161** (±0.072) | **0.7612** (±0.009) | **0.8475** (±0.034) | 0.0368 (±0.044) | **0.1586** (±0.084) | **0.1574** (±0.097) | **0.1375** (±0.092) | **0.1343** (±0.050) |

Brain = brain/glioma; Pros. = prostate. Bold/underline mark best/second among common-target results. § GeneQuery uses its native target and is excluded from ranking; † TissueFormer uses training-reference retrieval. Reference access is specified in Table A7.

# M Additional results

## M.1 Supplementary benchmark tables

Supplementary Table S1 reports macro-averaged results on the high-variance, HVG, and low-variance panels, and Supplementary Table S2 gives their per-cancer PCC values. Supplementary Table S3 reports the full 200-gene held-out evaluation and per-cancer results for the six-gene immune subset. Supplementary Figure S1 extends the EPCAM comparison to nine additional baselines. The TENX53 examples in Section M.2 use shared within-gene color ranges to complement the main cross-cancer localization analysis. Metric definitions and aggregation follow Section 4.

Table S1: **Prediction performance on three additional supervised gene panels.** Panels (A)–(C) report high-variance, highly variable (HVG), and low-variance genes. Values are five-cancer macro means, with across-cancer standard deviations beneath them.

| | (A) High-variance genes | | | | | (B) Highly variable genes (HVG) | | | | | (C) Low-variance genes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | $\text{PCC}_{\text{spot}}\uparrow$ | $\rho_{\text{spot}}\uparrow$ | $\text{PCC}_{\text{gene}}\uparrow$ | $\rho_{\text{gene}}\uparrow$ | RMSE↓ | $\text{PCC}_{\text{spot}}\uparrow$ | $\rho_{\text{spot}}\uparrow$ | $\text{PCC}_{\text{gene}}\uparrow$ | $\rho_{\text{gene}}\uparrow$ | RMSE↓ | $\text{PCC}_{\text{spot}}\uparrow$ | $\rho_{\text{spot}}\uparrow$ | $\text{PCC}_{\text{gene}}\uparrow$ | $\rho_{\text{gene}}\uparrow$ | RMSE↓ |
| HE2Gene | 0.3935 | 0.3307 | 0.0756 | 0.0729 | 0.8953 | 0.5072 | 0.4074 | 0.0503 | 0.0518 | 0.7014 | 0.0558 | 0.0457 | 0.0178 | 0.0291 | 0.2735 |
| | (±0.140) | (±0.119) | (±0.059) | (±0.060) | (±0.076) | (±0.086) | (±0.088) | (±0.043) | (±0.049) | (±0.082) | (±0.073) | (±0.067) | (±0.019) | (±0.033) | (±0.065) |
| ST-Net | 0.3743 | 0.2992 | 0.0276 | 0.0318 | 0.9492 | 0.4895 | 0.3833 | 0.0236 | 0.0257 | 0.7449 | 0.0294 | 0.0265 | 0.0085 | 0.0147 | 0.3294 |
| | (±0.134) | (±0.110) | (±0.042) | (±0.042) | (±0.182) | (±0.137) | (±0.125) | (±0.028) | (±0.031) | (±0.146) | (±0.030) | (±0.030) | (±0.018) | (±0.037) | (±0.110) |
| HisToGene | 0.4315 | 0.3587 | 0.0714 | 0.0655 | 0.8424 | 0.5530 | 0.4594 | 0.0377 | 0.0434 | 0.6722 | 0.0631 | 0.0445 | 0.0263 | 0.0539 | 0.2673 |
| | (±0.110) | (±0.101) | (±0.048) | (±0.049) | (±0.070) | (±0.083) | (±0.089) | (±0.032) | (±0.037) | (±0.065) | (±0.065) | (±0.038) | (±0.022) | (±0.038) | (±0.066) |
| Hist2ST | 0.3992 | 0.3234 | 0.0018 | 0.0019 | 0.8672 | 0.5218 | 0.4249 | 0.0010 | 0.0009 | 0.6908 | 0.0559 | 0.0451 | -0.0001 | -0.0006 | 0.2717 |
| | (±0.124) | (±0.103) | (±0.001) | (±0.001) | (±0.063) | (±0.104) | (±0.103) | (±0.001) | (±0.001) | (±0.067) | (±0.062) | (±0.061) | (±0.001) | (±0.001) | (±0.063) |
| THItoGene | 0.3964 | 0.3271 | 0.0011 | 0.0010 | 0.8615 | 0.5209 | 0.4258 | 0.0005 | 0.0006 | 0.6889 | 0.0648 | 0.0511 | 0.0004 | 0.0006 | 0.2721 |
| | (±0.127) | (±0.108) | (±0.003) | (±0.003) | (±0.071) | (±0.100) | (±0.107) | (±0.001) | (±0.001) | (±0.069) | (±0.065) | (±0.060) | (±0.002) | (±0.002) | (±0.067) |
| BrSTNet | 0.3787 | 0.3047 | -0.0105 | -0.0079 | 0.8858 | 0.5014 | 0.3984 | -0.0031 | -0.0066 | 0.7045 | 0.0436 | 0.0374 | -0.0101 | -0.0187 | 0.2857 |
| | (±0.131) | (±0.108) | (±0.046) | (±0.046) | (±0.079) | (±0.117) | (±0.117) | (±0.031) | (±0.037) | (±0.077) | (±0.053) | (±0.052) | (±0.024) | (±0.046) | (±0.062) |
| DeepSpot-M | 0.3869 | 0.3155 | 0.0674 | 0.0581 | 0.8877 | 0.5156 | 0.4223 | 0.0362 | 0.0501 | 0.6987 | 0.0583 | 0.0428 | 0.0343 | 0.0731 | 0.2720 |
| | (±0.120) | (±0.102) | (±0.021) | (±0.017) | (±0.067) | (±0.103) | (±0.105) | (±0.022) | (±0.023) | (±0.079) | (±0.057) | (±0.044) | (±0.027) | (±0.045) | (±0.062) |
| STFlow | **0.5236** | **0.4541** | **0.1196** | **0.1109** | **0.7835** | **0.6384** | **0.5199** | **0.0887** | 0.0892 | **0.6051** | **0.0960** | **0.0696** | 0.0260 | 0.0434 | **0.2654** |
| | (±0.104) | (±0.096) | (±0.034) | (±0.029) | (±0.061) | (±0.071) | (±0.090) | (±0.030) | (±0.029) | (±0.064) | (±0.065) | (±0.045) | (±0.015) | (±0.024) | (±0.066) |
| TCGN | 0.3557 | 0.2891 | 0.0451 | 0.0501 | 0.9551 | 0.4672 | 0.3701 | 0.0372 | 0.0437 | 0.7450 | 0.0318 | 0.0333 | 0.0203 | 0.0431 | 0.2765 |
| | (±0.142) | (±0.117) | (±0.042) | (±0.043) | (±0.090) | (±0.146) | (±0.129) | (±0.030) | (±0.035) | (±0.094) | (±0.042) | (±0.042) | (±0.028) | (±0.048) | (±0.064) |
| GeneQuery§ | – | – | 0.0527 | 0.0589 | – | – | – | 0.0371 | 0.0425 | – | – | – | 0.0297 | 0.0326 | – |
| | | | (±0.045) | (±0.061) | | | | (±0.032) | (±0.041) | | | | (±0.025) | (±0.031) | |
| SGN | 0.3372 | 0.2493 | 0.0507 | 0.0413 | 0.9483 | 0.4382 | 0.2766 | 0.0208 | 0.0313 | 0.7280 | 0.0071 | 0.0070 | 0.0273 | 0.0580 | 0.3136 |
| | (±0.119) | (±0.086) | (±0.032) | (±0.025) | (±0.070) | (±0.110) | (±0.072) | (±0.009) | (±0.009) | (±0.082) | (±0.007) | (±0.007) | (±0.017) | (±0.029) | (±0.044) |
| TissueFormer† | 0.4179 | 0.3515 | 0.0737 | 0.0674 | 0.8817 | 0.5188 | 0.4249 | 0.0485 | 0.0527 | 0.7128 | 0.0650 | 0.0456 | 0.0191 | 0.0372 | 0.2782 |
| | (±0.137) | (±0.118) | (±0.036) | (±0.030) | (±0.084) | (±0.112) | (±0.125) | (±0.030) | (±0.030) | (±0.039) | (±0.050) | (±0.031) | (±0.012) | (±0.022) | (±0.064) |
| GeneRoute | 0.4909 | 0.4128 | 0.1183 | 0.1068 | 0.8319 | 0.6033 | 0.4673 | 0.0850 | **0.0938** | 0.6450 | 0.0358 | 0.0388 | **0.0372** | **0.0773** | 0.2866 |
| | (±0.101) | (±0.095) | (±0.051) | (±0.054) | (±0.064) | (±0.066) | (±0.090) | (±0.035) | (±0.038) | (±0.056) | (±0.037) | (±0.038) | (±0.024) | (±0.044) | (±0.062) |

*Notes.* Standard deviations are calculated across five cancer means ($n = 5$). Bold and underline mark the best and second-best common-target means; GeneQuery is shown as a native-target reference and excluded from ranking. § GeneQuery uses its native section-wise target; spot-wise metrics and RMSE are omitted from the common-target comparison. † TissueFormer uses in-context retrieval; SGN is trained directly on the common target. Methods use common benchmark partitions and scored panels; information access is detailed in Supplementary Table A7. Comparison conditions follow Section 4.

Table S2: **Per-cancer prediction performance on supervised gene panels.** Spot-wise and gene-wise PCC are reported for high-variance, highly variable (HVG), and low-variance genes. Values are section means within each cancer, with across-section standard deviations beneath them.

**(A) High-variance genes**

| | $PCC_{spot}$ | | | | | $PCC_{gene}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | Brain | Breast | CRC | RCC | Pros. | Brain | Breast | CRC | RCC | Pros. |
| HE2Gene | 0.1821 (±0.062) | 0.5214 (±0.048) | 0.4161 (±0.066) | 0.3369 (±0.116) | 0.5107 (±0.071) | 0.0360 (±0.048) | 0.0421 (±0.062) | 0.0842 (±0.043) | 0.0406 (±0.040) | <u>0.1751</u> (±0.025) |
| ST-Net | 0.1466 (±0.038) | 0.5020 (±0.057) | 0.4123 (±0.100) | 0.3926 (±0.063) | 0.4179 (±0.117) | 0.0120 (±0.038) | -0.0180 (±0.101) | 0.0445 (±0.053) | 0.0081 (±0.033) | 0.0913 (±0.048) |
| HisToGene | 0.2517 (±0.129) | 0.5427 (±0.034) | 0.4870 (±0.095) | 0.4246 (±0.053) | 0.4517 (±0.173) | 0.0363 (±0.087) | 0.0995 (±0.040) | 0.0585 (±0.033) | 0.0223 (±0.035) | 0.1401 (±0.064) |
| Hist2ST | 0.1995 (±0.082) | 0.5391 (±0.074) | 0.4081 (±0.061) | 0.4080 (±0.028) | 0.4413 (±0.134) | 0.0005 (±0.001) | 0.0033 (±0.003) | 0.0023 (±0.013) | 0.0011 (±0.002) | 0.0017 (±0.002) |
| THItoGene | 0.1889 (±0.069) | 0.5323 (±0.053) | 0.4150 (±0.099) | 0.3935 (±0.051) | 0.4523 (±0.082) | -0.0035 (±0.002) | 0.0051 (±0.008) | 0.0033 (±0.012) | 0.0005 (±0.003) | 0.0002 (±0.001) |
| BrSTNet | 0.1572 (±0.052) | 0.5011 (±0.069) | 0.4054 (±0.092) | 0.3948 (±0.071) | 0.4352 (±0.099) | -0.0044 (±0.034) | 0.0176 (±0.114) | 0.0403 (±0.047) | -0.0268 (±0.039) | -0.0794 (±0.028) |
| DeepSpot-M | 0.2019 (±0.065) | 0.5335 (±0.054) | 0.4016 (±0.118) | 0.3720 (±0.030) | 0.4254 (±0.144) | 0.0464 (±0.102) | 0.0979 (±0.017) | 0.0592 (±0.044) | 0.0785 (±0.015) | 0.0550 (±0.023) |
| STFlow | **0.3752** (±0.244) | **0.6356** (±0.000) | **0.5346** (±0.116) | **0.4726** (±0.017) | **0.5999** (±0.164) | **0.0857** (±0.076) | <u>0.1480</u> (±0.042) | **0.1189** (±0.063) | <u>0.0865</u> (±0.062) | 0.1590 (±0.075) |
| TCGN | 0.1602 (±0.060) | 0.5235 (±0.101) | 0.2697 (±0.077) | 0.4242 (±0.016) | 0.4009 (±0.126) | 0.0352 (±0.042) | 0.0464 (±0.003) | 0.0136 (±0.022) | 0.0137 (±0.007) | 0.1164 (±0.066) |
| GeneQuery§ | – | – | – | – | – | -0.0097 (±0.027) | 0.0200 (±0.081) | 0.0811 (±0.106) | 0.0936 (±0.029) | 0.0785 (±0.088) |
| SGN | 0.1696 (±0.071) | 0.4970 (±0.052) | 0.3256 (±0.116) | 0.3120 (±0.028) | 0.3819 (±0.144) | <u>0.0697</u> (±0.095) | 0.0979 (±0.004) | 0.0291 (±0.040) | 0.0367 (±0.037) | 0.0204 (±0.038) |
| TissueFormer† | 0.2060 (±0.018) | 0.5467 (±0.115) | 0.4253 (±0.076) | 0.3814 (±0.032) | 0.5302 (±0.147) | 0.0251 (±0.022) | 0.1118 (±0.026) | 0.0916 (±0.044) | 0.0473 (±0.037) | 0.0927 (±0.036) |
| GeneRoute | <u>0.3408</u> (±0.146) | <u>0.5955</u> (±0.006) | <u>0.4889</u> (±0.115) | <u>0.4600</u> (±0.026) | <u>0.5695</u> (±0.146) | 0.0483 (±0.077) | **0.1499** (±0.045) | <u>0.1081</u> (±0.065) | **0.1030** (±0.055) | **0.1822** (±0.042) |

*Notes.* Standard deviations describe test-section variation within each cancer. Brain = brain/glioma; Pros. = prostate. Bold/underline mark the best and second-best common-target means in each column; GeneQuery is shown as a native-target reference and excluded from ranking. § GeneQuery uses its native section-wise target; † TissueFormer uses in-context retrieval. Reference access is detailed in Supplementary Table A7.

**Table S2 (continued)**
**(B) Highly variable genes (HVG)**

| | $PCC_{spot}$ | | | | | $PCC_{gene}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | Brain | Breast | CRC | RCC | Pros. | Brain | Breast | CRC | RCC | Pros. |
| HE2Gene | 0.3712 (±0.109) | 0.6068 (±0.023) | 0.5242 (±0.070) | 0.5358 (±0.058) | 0.4981 (±0.038) | 0.0075 (±0.018) | 0.0250 (±0.039) | 0.0771 (±0.039) | 0.0295 (±0.043) | 0.1122 (±0.011) |
| ST-Net | 0.3353 (±0.090) | 0.6253 (±0.045) | 0.5069 (±0.092) | 0.6175 (±0.035) | 0.3627 (±0.144) | 0.0030 (±0.008) | 0.0030 (±0.033) | 0.0510 (±0.047) | 0.0034 (±0.026) | 0.0577 (±0.032) |
| HisToGene | 0.4354 (±0.181) | 0.6312 (±0.034) | 0.5601 (±0.078) | 0.6281 (±0.017) | 0.5100 (±0.080) | 0.0090 (±0.022) | 0.0427 (±0.016) | 0.0463 (±0.027) | 0.0062 (±0.015) | 0.0842 (±0.038) |
| Hist2ST | 0.3884 (±0.141) | 0.6260 (±0.060) | 0.4841 (±0.058) | 0.6288 (±0.029) | 0.4819 (±0.060) | 0.0002 (±0.001) | 0.0003 (±0.002) | 0.0025 (±0.013) | 0.0000 (±0.001) | 0.0019 (±0.001) |
| THItoGene | 0.3811 (±0.128) | 0.6281 (±0.038) | 0.5005 (±0.085) | 0.6070 (±0.030) | 0.4877 (±0.023) | 0.0003 (±0.002) | -0.0003 (±0.000) | 0.0012 (±0.004) | -0.0008 (±0.001) | 0.0021 (±0.003) |
| BrSTNet | 0.3475 (±0.110) | 0.6174 (±0.060) | 0.5016 (±0.084) | 0.6112 (±0.023) | 0.4293 (±0.059) | 0.0041 (±0.010) | 0.0095 (±0.042) | 0.0344 (±0.029) | -0.0165 (±0.011) | -0.0471 (±0.020) |
| DeepSpot-M | 0.3919 (±0.122) | 0.6300 (±0.039) | 0.4909 (±0.090) | 0.6123 (±0.029) | 0.4532 (±0.073) | 0.0078 (±0.024) | 0.0191 (±0.009) | 0.0499 (±0.021) | 0.0456 (±0.001) | 0.0588 (±0.029) |
| STFlow | **0.5358** (±0.268) | **0.7238** (±0.001) | **0.6052** (±0.100) | **0.6655** (±0.036) | **0.6617** (±0.099) | **0.0486** (±0.045) | **0.0854** (±0.051) | **0.1197** (±0.072) | 0.0726 (±0.050) | 0.1173 (±0.057) |
| TCGN | 0.3296 (±0.113) | 0.6181 (±0.085) | 0.3704 (±0.078) | 0.6330 (±0.024) | 0.3850 (±0.079) | 0.0081 (±0.008) | 0.0477 (±0.008) | 0.0334 (±0.021) | 0.0142 (±0.012) | 0.0826 (±0.045) |
| GeneQuery[§] | – | – | – | – | – | 0.0011 (±0.013) | 0.0043 (±0.034) | 0.0509 (±0.058) | 0.0634 (±0.015) | 0.0657 (±0.056) |
| SGN | 0.3400 (±0.120) | 0.5735 (±0.026) | 0.3899 (±0.098) | 0.5388 (±0.041) | 0.3490 (±0.106) | 0.0191 (±0.030) | 0.0144 (±0.004) | 0.0143 (±0.026) | 0.0194 (±0.007) | 0.0366 (±0.024) |
| TissueFormer[†] | 0.3387 (±0.044) | 0.6027 (±0.125) | 0.4807 (±0.079) | 0.5908 (±0.020) | 0.5813 (±0.092) | 0.0054 (±0.005) | 0.0452 (±0.001) | 0.0793 (±0.031) | 0.0391 (±0.022) | 0.0735 (±0.026) |
| GeneRoute | 0.5107 (±0.185) | 0.6736 (±0.002) | 0.5615 (±0.093) | 0.6490 (±0.029) | 0.6214 (±0.056) | 0.0274 (±0.030) | 0.0852 (±0.037) | 0.1089 (±0.054) | **0.0849** (±0.044) | **0.1185** (±0.046) |

*Notes.* Standard deviations describe test-section variation within each cancer. Brain = brain/glioma; Pros. = prostate. Bold/underline mark the best and second-best common-target means in each column; GeneQuery is shown as a native-target reference and excluded from ranking. § GeneQuery uses its native section-wise target; † TissueFormer uses in-context retrieval. Reference access is detailed in Supplementary Table A7.

**Table S2 (continued)**
**(C) Low-variance genes**

| | $PCC_{spot}$ | | | | | $PCC_{gene}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | Brain | Breast | CRC | RCC | Pros. | Brain | Breast | CRC | RCC | Pros. |
| HE2Gene | -0.0001 (±0.010) | 0.0146 (±0.020) | 0.0850 (±0.075) | 0.0079 (±0.024) | **0.1714** (±0.119) | 0.0061 (±0.009) | 0.0056 (±0.012) | 0.0180 (±0.012) | 0.0080 (±0.005) | 0.0511 (±0.016) |
| ST-Net | -0.0056 (±0.001) | 0.0188 (±0.019) | 0.0394 (±0.029) | 0.0196 (±0.009) | 0.0749 (±0.074) | 0.0016 (±0.005) | -0.0081 (±0.005) | 0.0030 (±0.011) | 0.0061 (±0.006) | 0.0399 (±0.048) |
| HisToGene | -0.0002 (±0.008) | 0.0194 (±0.019) | <u>0.1614</u> (±0.144) | 0.0421 (±0.026) | 0.0931 (±0.089) | 0.0095 (±0.020) | 0.0294 (±0.012) | 0.0225 (±0.011) | 0.0077 (±0.009) | 0.0624 (±0.021) |
| Hist2ST | -0.0073 (±0.007) | 0.0237 (±0.027) | 0.0817 (±0.069) | 0.0293 (±0.016) | 0.1519 (±0.121) | -0.0012 (±0.001) | 0.0012 (±0.001) | -0.0001 (±0.002) | -0.0000 (±0.000) | -0.0003 (±0.001) |
| THItoGene | -0.0061 (±0.003) | 0.0275 (±0.028) | 0.1030 (±0.092) | <u>0.0430</u> (±0.023) | <u>0.1565</u> (±0.124) | 0.0012 (±0.001) | 0.0025 (±0.004) | -0.0003 (±0.001) | 0.0006 (±0.001) | -0.0019 (±0.000) |
| BrSTNet | -0.0066 (±0.001) | 0.0219 (±0.023) | 0.0404 (±0.027) | 0.0297 (±0.013) | 0.1328 (±0.097) | -0.0012 (±0.005) | 0.0033 (±0.014) | 0.0044 (±0.006) | -0.0049 (±0.004) | -0.0523 (±0.016) |
| DeepSpot-M | -0.0014 (±0.007) | 0.0291 (±0.028) | 0.1188 (±0.111) | 0.0244 (±0.014) | 0.1205 (±0.089) | 0.0061 (±0.023) | 0.0336 (±0.013) | <u>0.0304</u> (±0.017) | <u>0.0232</u> (±0.008) | **0.0784** (±0.028) |
| STFlow | **0.0329** (±0.032) | **0.0495** (±0.014) | **0.1800** (±0.147) | **0.0672** (±0.032) | 0.1504 (±0.081) | <u>0.0121</u> (±0.011) | 0.0243 (±0.009) | 0.0248 (±0.014) | 0.0166 (±0.012) | 0.0519 (±0.033) |
| TCGN | -0.0083 (±0.015) | 0.0261 (±0.028) | 0.0161 (±0.025) | 0.0222 (±0.018) | 0.1028 (±0.081) | 0.0088 (±0.016) | 0.0158 (±0.008) | 0.0004 (±0.006) | 0.0079 (±0.002) | 0.0686 (±0.028) |
| GeneQuery[§] | – | – | – | – | – | 0.0045 (±0.010) | 0.0072 (±0.012) | 0.0381 (±0.055) | 0.0322 (±0.010) | 0.0666 (±0.038) |
| SGN | -0.0040 (±0.006) | 0.0049 (±0.001) | 0.0130 (±0.016) | 0.0106 (±0.005) | 0.0111 (±0.007) | **0.0137** (±0.024) | <u>0.0342</u> (±0.002) | 0.0173 (±0.020) | 0.0181 (±0.015) | 0.0533 (±0.030) |
| TissueFormer[†] | <u>0.0178</u> (±0.027) | <u>0.0323</u> (±0.023) | 0.1394 (±0.096) | 0.0423 (±0.031) | 0.0934 (±0.034) | 0.0061 (±0.004) | 0.0209 (±0.010) | 0.0255 (±0.013) | 0.0081 (±0.004) | 0.0351 (±0.020) |
| GeneRoute | 0.0053 (±0.004) | 0.0224 (±0.028) | 0.0579 (±0.054) | 0.0043 (±0.009) | 0.0891 (±0.061) | 0.0087 (±0.021) | **0.0400** (±0.008) | **0.0354** (±0.024) | **0.0266** (±0.013) | <u>0.0755</u> (±0.041) |

*Notes.* Standard deviations describe test-section variation within each cancer. Brain = brain/glioma; Pros. = prostate. Bold/underline mark the best and second-best common-target means in each column; GeneQuery is shown as a native-target reference and excluded from ranking. § GeneQuery uses its native section-wise target; † TissueFormer uses in-context retrieval. Reference access is detailed in Supplementary Table A7.

Table S3: **Held-out gene prediction across five cancer contexts.** The upper panel reports all 200 genes: five-cancer macro Gene-PCC, Gene-Spearman, and RMSE, followed by per-cancer Gene-PCC. The lower panel reports per-cancer Gene-PCC for the six-gene immune-receptor subset, whose macro means appear in Table 2. Spot-level metrics are omitted from the held-out displays.

**All 200 held-out genes**

| | Five-cancer macro | | | Per-cancer $\mathrm{PCC}_{\mathrm{gene}}$ | | | | |
|---|---|---|---|---|---|---|---|---|
| **Method** | $\mathrm{PCC}_{\mathrm{gene}}\uparrow$ | $\rho_{\mathrm{gene}}\uparrow$ | RMSE$\downarrow$ | Brain | Breast | CRC | RCC | Pros. |
| DeepSpot-M | 0.0167 (±0.008) | <u>0.0402</u> (±0.020) | <u>0.4410</u> (±0.130) | 0.0048 (±0.020) | 0.0209 (±0.006) | 0.0150 (±0.010) | <u>0.0176</u> (±0.006) | 0.0252 (±0.014) |
| TissueFormer† | <u>0.0184</u> (±0.008) | 0.0304 (±0.014) | **0.3768** (±0.044) | 0.0063 (±0.005) | <u>0.0274</u> (±0.021) | <u>0.0231</u> (±0.007) | 0.0154 (±0.011) | 0.0200 (±0.007) |
| SGN‡ | 0.0103 (±0.011) | 0.0269 (±0.023) | 0.4448 (±0.129) | <u>0.0097</u> (±0.008) | 0.0086 (±0.007) | -0.0005 (±0.017) | 0.0049 (±0.009) | <u>0.0289</u> (±0.010) |
| GeneQuery§ | 0.0197 (±0.017) | 0.0217 (±0.022) | – | 0.0028 (±0.013) | 0.0000 (±0.018) | 0.0262 (±0.043) | 0.0351 (±0.009) | 0.0344 (±0.026) |
| GeneRoute | **0.0273** (±0.010) | **0.0515** (±0.023) | 0.4535 (±0.114) | **0.0110** (±0.020) | **0.0365** (±0.004) | **0.0289** (±0.021) | **0.0273** (±0.012) | **0.0331** (±0.014) |

**Six-gene immune-receptor subset**

| | Per-cancer $\mathrm{PCC}_{\mathrm{gene}}$ | | | | |
|---|---|---|---|---|---|
| **Method** | Brain | Breast | CRC | RCC | Pros. |
| DeepSpot-M | <u>0.0202</u> (±0.022) | -0.1457 (±0.193) | -0.0845 (±0.077) | -0.0674 (±0.124) | -0.0988 (±0.025) |
| TissueFormer† | -0.0099 (±0.011) | **0.1121** (±0.101) | <u>0.0451</u> (±0.050) | <u>0.0752</u> (±0.075) | <u>0.0479</u> (±0.045) |
| SGN‡ | -0.0084 (±0.014) | -0.2002 (±0.166) | -0.1251 (±0.075) | -0.0821 (±0.215) | -0.0297 (±0.042) |
| GeneQuery§ | 0.0057 (±0.016) | -0.0583 (±0.016) | -0.0134 (±0.057) | 0.1166 (±0.031) | 0.0072 (±0.062) |
| GeneRoute | **0.0319** (±0.015) | <u>0.0914</u> (±0.099) | **0.1018** (±0.107) | **0.1156** (±0.083) | **0.1592** (±0.043) |

*Notes.* The comparison includes GENEROUTE and four baselines with held-out predictions. Macro columns carry the sample standard deviation across the five cancer means; per-cancer columns carry the sample standard deviation across test sections within each cancer. TissueFormer uses in-context retrieval, ‡ SGN uses the training-only calibration in Eq. (36), and GeneQuery uses a native target; dashes indicate omitted common-scale metrics (GeneQuery's RMSE). Information access is detailed in Supplementary Table A7. Bold and underline rank common-target means only; GeneQuery is a native-target reference outside this ranking. The comparison retains the stated reference-access conditions. Exclusion and checkpoint-selection conditions follow Sections 4 and J.

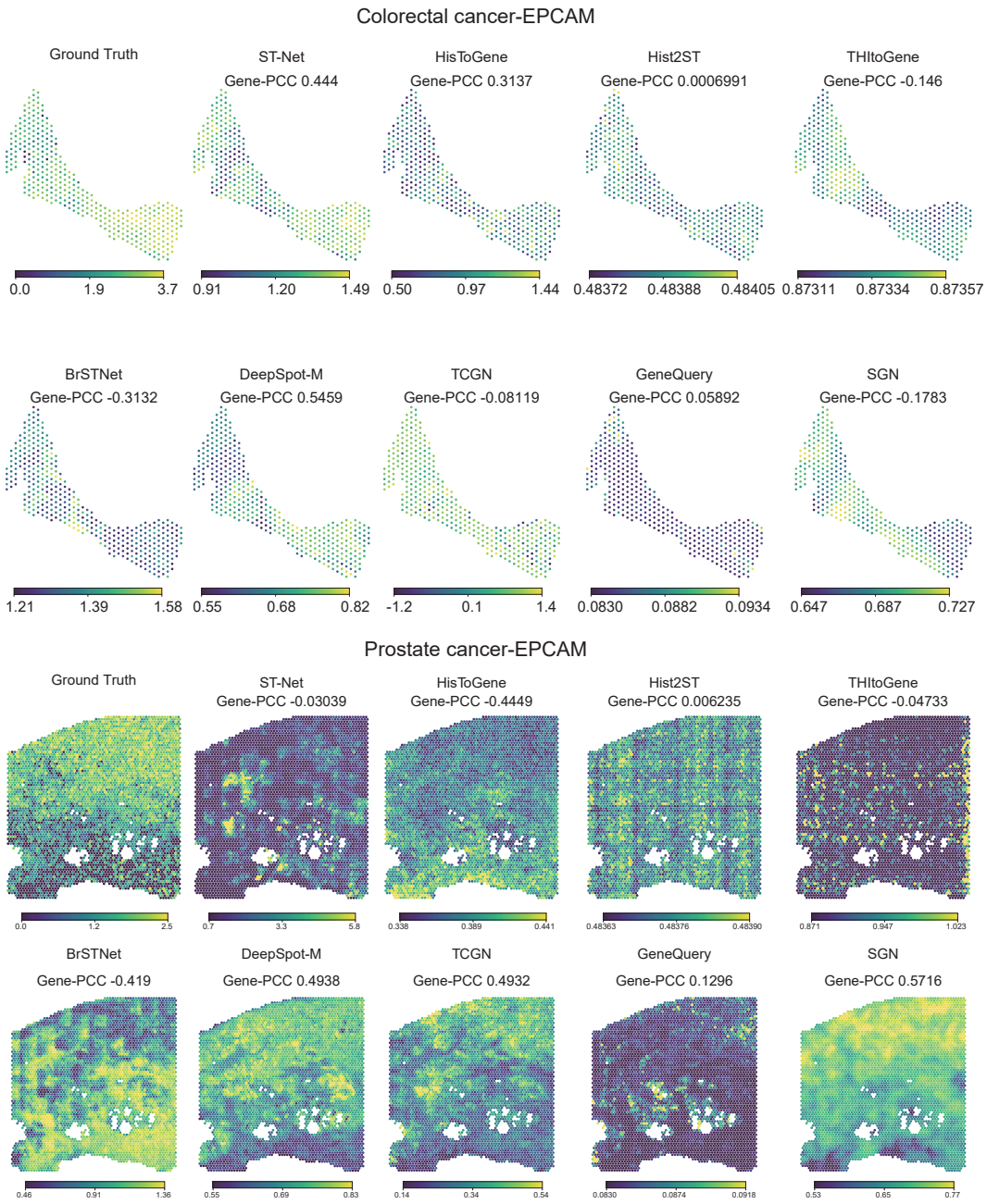


Figure S1: **Additional EPCAM predictions in colorectal and prostate cancer.** Measured expression is compared with ST-Net, HisToGene, Hist2ST, THItoGene, BrSTNet, DeepSpot-M, TCGN, GeneQuery, and SGN on the sections shown in Fig. 2. Each prediction is annotated with within-section gene-wise PCC. Maps use separate color limits; GeneQuery retains its native target. With the main figure, these examples cover all 12 comparator methods on the two sections.

### M.2 BREAST-CANCER MARKERS AND CELL-ABUNDANCE CONTEXT IN TENX53

**Marker expression on a common display scale.** Whereas the main EPCAM analysis compares epithelial localization across cancers, the TENX53 examples examine several genes within one breast-cancer section. ESR1, ERBB2, and AR are displayed using the common full-library expression transformation and a fixed display range for each gene. The same range is applied across common-target methods, without method-specific rescaling or smoothing. GeneQuery is an explicitly marked exception because it retains its native section-wise 0–1 target based on $\log(1 + x)$ raw counts. These panels therefore complement the separately scaled EPCAM maps by showing spatial patterns and expression contrast on a shared within-gene display scale.

The breast-cancer sample analyzed as TENX53 is an invasive ductal carcinoma with Grade III pathology, T2N0M0 staging, ER-positive/PR-negative status, and a HercepTest 2+ IHC record (10x Genomics, 2022). The latter is equivocal and is not interpreted as HER2 positivity (Wolff et al., 2023). GENEROUTE achieves gene-wise PCC values of 0.417, 0.280, and 0.209 for ESR1, ERBB2, and AR, respectively. These are spatial-expression examples, not a validation of clinical receptor classification.

**Inferred cell-abundance context.** Cell2location posterior mean abundances are aligned to all 4,898 spots in TENX53 and provide nine broad cell classes (Kleshchevnikov et al., 2022) (Fig. S5). These abundance estimates provide additional cellular context; they are neither manual pathology labels nor independent measurements of the predicted genes.

EPCAM is nonzero at 2,853 spots (58.25%) and belongs to the supervised rather than the permanently held-out gene set. The abundance maps provide descriptive cellular context and do not independently validate EPCAM prediction.

## N ADDITIONAL ABLATIONS AND ROBUSTNESS

**Retrained component variants.** Table 3 compares the complete GeneRoute model with 12 retrained removal conditions, all using the same training budget of 20 epochs of stage 1. The interventions remove priority-guided gene sampling; restrict images to $\mathrm{R0}$; replace ESM2 and RNA-FM inputs with random sequence embeddings; zero GenePT embeddings; remove DepMap or STRING; remove Cancer Cards; remove JUMP ORF/CRISPR programs; remove GO/Reactome/Hallmark programs; remove all program-prior keys and values; remove decoder cancer modulation; or remove all prior sources jointly. The complete model is the reference for the percentage changes.

We randomly selected four test sections per cancer from the 33-section common test pool, for 20 sections total. The complete model and all perturbations use the same training partition, selected test sections, four reporting panels, target transformation, and masked gene-PCC definition.

**Fixed-weight sensitivity analysis.** An auxiliary analysis changes inputs to the earlier stage-1 checkpoint without retraining. It holds fixed the evaluation sections, 6,088-gene query axis, four reporting panels, target transformation, and availability masks across interventions. Disabling Cancer Cards changes gene-wise PCC by at most $3.99 \times 10^{-5}$. Disabling program-prior contributions reduces PCC in 17 of 20 cancer–panel combinations, including all five HVG combinations, while retaining sequence/text and graph inputs. Restricting readout to $\mathrm{R0}$ improves PCC in 5 of 20 combinations; prostate high-variance PCC decreases from 0.15777 to 0.10856. These interventions describe utilization by the earlier checkpoint, not the contribution of components after retraining. Removing cards at inference also does not undo priority sampling during training.

**Scope of the ablations.** The fixed-weight sensitivity analysis and retrained ablation address different questions and should not be interpreted interchangeably.

**Controlled queries.** The conditioning interface permits changing the requested gene or cancer package while holding histology fixed. Such a query produces a conditional prediction for the new input; it does not simulate a knockout, overexpression, or other causal perturbation. The present model therefore provides conditional expression estimates rather than validated causal effects.

Table S4: **Training and test composition.** All methods use these sections under the common benchmark partition. The aggregate validation partition contains 23 sections and 82,484 spots; the rows below report training and test composition. Train, validation, and test partitions are disjoint at both the patient and specimen levels.

| Cancer context | Train sections | Train spots | Test sections | Test spots |
|---|---|---|---|---|
| Brain/glioma | 15 | 42,988 | 8 | 22,684 |
| Breast | 14 | 33,665 | 6 | 16,955 |
| Colorectal | 38 | 122,268 | 8 | 12,184 |
| Kidney RCC | 17 | 54,013 | 4 | 13,530 |
| Prostate | 24 | 65,524 | 7 | 28,696 |
| Total | 108 | 318,458 | 33 | 94,049 |

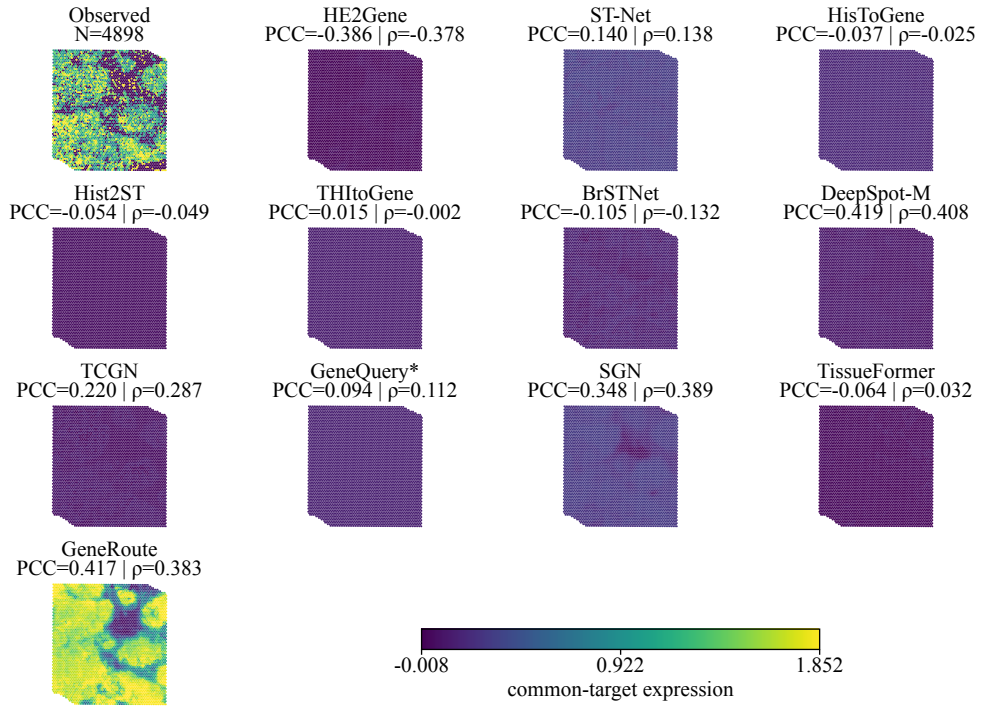


Figure S2: **ESR1 expression in breast-cancer section TENX53.** Measured expression and predictions share a fixed color range on the common log-normalized target, except for GeneQuery, whose native section-wise 0–1 target based on $\log(1 + x)$ raw counts is marked by an asterisk. No method-specific rescaling or smoothing is applied to common-target predictions. TissueFormer uses in-context retrieval. GeneRoute achieves gene-wise PCC $= 0.417$ and Spearman's $\rho = 0.383$.

## O Compute and efficiency

The computational pipeline comprises prior preparation, frozen image-feature extraction, trainable model optimization, and inference over cached tokens. Caching avoids repeated image encoding, while gene chunking limits attention memory. Runtime still increases with the number of queried genes.

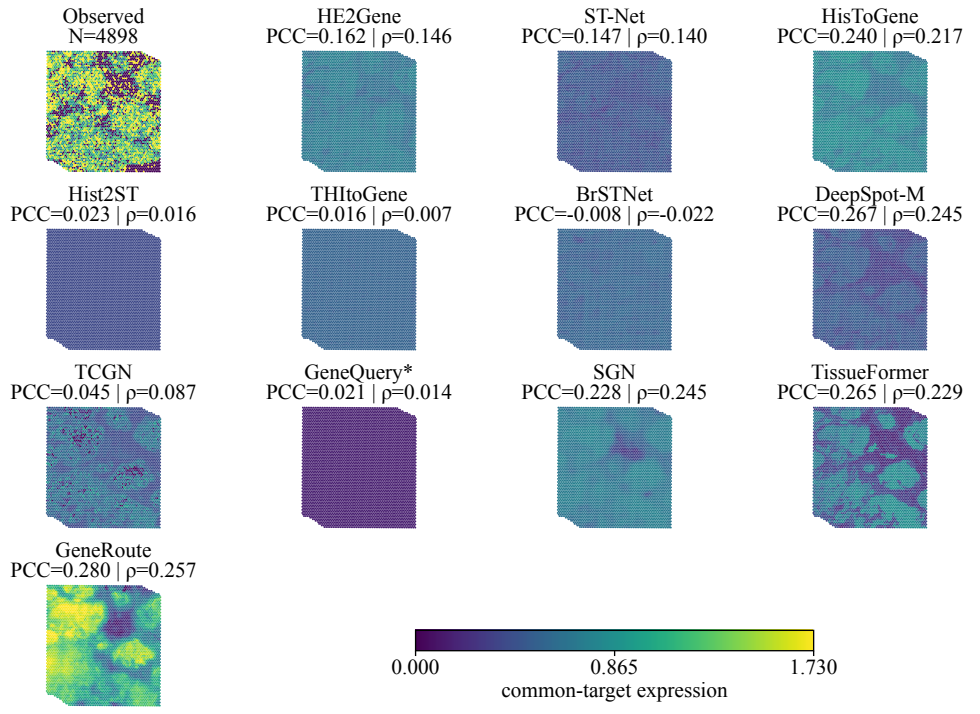


Figure S3: **ERBB2 expression in breast-cancer section TENX53.** Display settings and the GeneQuery exception follow Fig. S2. GENEROUTE achieves gene-wise PCC $= 0.280$ and Spearman's $\rho = 0.257$. The sample's HercepTest 2+ result is equivocal; this expression map does not establish clinical HER2 positivity (Wolff et al., 2023).

# P REPRODUCIBILITY MATERIALS

Table S4 gives the training, validation, and test composition. The architecture, source preprocessing, and training procedure are specified in Sections A, B, D, E, G and K. Target transformations, reference access, and metric aggregation are defined in Sections L and L.1.

The supplementary files `u200_genes.csv` and `supervised_panel_membership.csv` provide the permanent held-out genes and supervised panel memberships; `gene_panel_audit.json` records their intersections. Selection strata and the deterministic seed are given in Section J, and the complete held-out list also appears in Section S. These 200 genes are excluded from all GeneRoute expression training and model selection.

# Q EXTENDED RELATED WORK

## Q.1 HISTOLOGY-TO-SPATIAL-EXPRESSION PREDICTION

Early work established the predictive relationship between tissue morphology and gene expression. HE2RNA studied this relationship at the whole-slide level, whereas ST-Net predicted spatially resolved expression from local histology (Schmauch et al., 2020; He et al., 2020). HE2Gene and BrSTNet further developed image-based expression prediction (Chen et al., 2024; Rahaman et al., 2023). Subsequent methods expanded the spatial context: HisToGene, Hist2ST, THItoGene, and TCGN model relationships among spots using attention or graph-based interactions (Pang et al., 2021; Zeng et al., 2022; Jia et al., 2024; Xiao et al., 2024). TRIPLEX and M2OST combine information

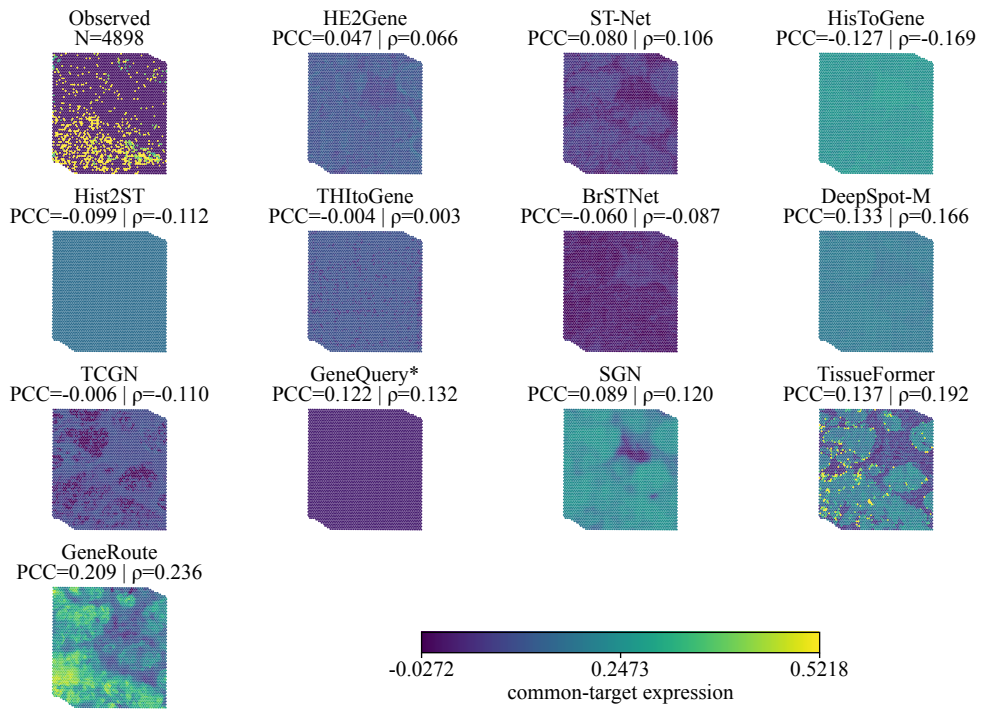


Figure S4: **AR expression in breast-cancer section TENX53.** Display settings and the GeneQuery exception follow Fig. S2. GENEROUTE achieves gene-wise PCC $= 0.209$ and Spearman's $\rho = 0.236$. AR provides an additional spatial-expression example rather than a receptor-status classification.

across image scales, while iStar focuses on finer-resolution expression reconstruction (Chung et al., 2024; Wang et al., 2025b; Zhang et al., 2024).

Other approaches emphasize the learning objective or the transferability of the representation. BLEEP and mclSTExp align histological and transcriptomic features through contrastive learning (Xie et al., 2023; Min et al., 2024); STFlow, STPath, and Stem explore generative formulations (Huang et al., 2025a;b; Zhu et al., 2025). RNAPath, OmiCLIP, GHIST, HESpotEx, sCellST, HistoPrism, and TissueFormer further extend the range of prediction settings (Cisternino et al., 2024; Chen et al., 2025; Fu et al., 2025; Yin et al., 2026; Chadoutaud et al., 2026; Hu et al., 2026; Wu et al., 2026). Resources such as HEST-1k and systematic benchmarks support comparisons across datasets and modeling paradigms (Jaume et al., 2024; Wang et al., 2025a). Much of this progress concerns visual encoding and spatial modeling; prediction beyond the supervised gene panel requires an additional mechanism for representing and querying genes.

### Q.2 GENE-AWARE SPATIAL EXPRESSION PREDICTION

Gene-aware models replace, or augment, fixed gene-indexed outputs with explicit molecular representations. GeneQuery introduces gene queries, and SGN uses functional and phenotype descriptions to support prediction for genes without spatial-expression supervision (Xiong et al., 2024; Yang et al., 2024). AGP-Net shares information through gene descriptions, CoMTIP jointly models gene, expression, and image information, and DeepSpot-M combines gene queries with a hypernetwork for transcriptome-scale prediction (Yang et al., 2025; Qian et al., 2025; Nonchev et al., 2026). These studies show that molecular priors can extend the set of genes accessible to a histology-based model.

GENEROUTE builds on this gene-conditioned formulation. Its focus is the interaction between gene information, cancer context, and multiscale morphology: shared gene representations are adapted

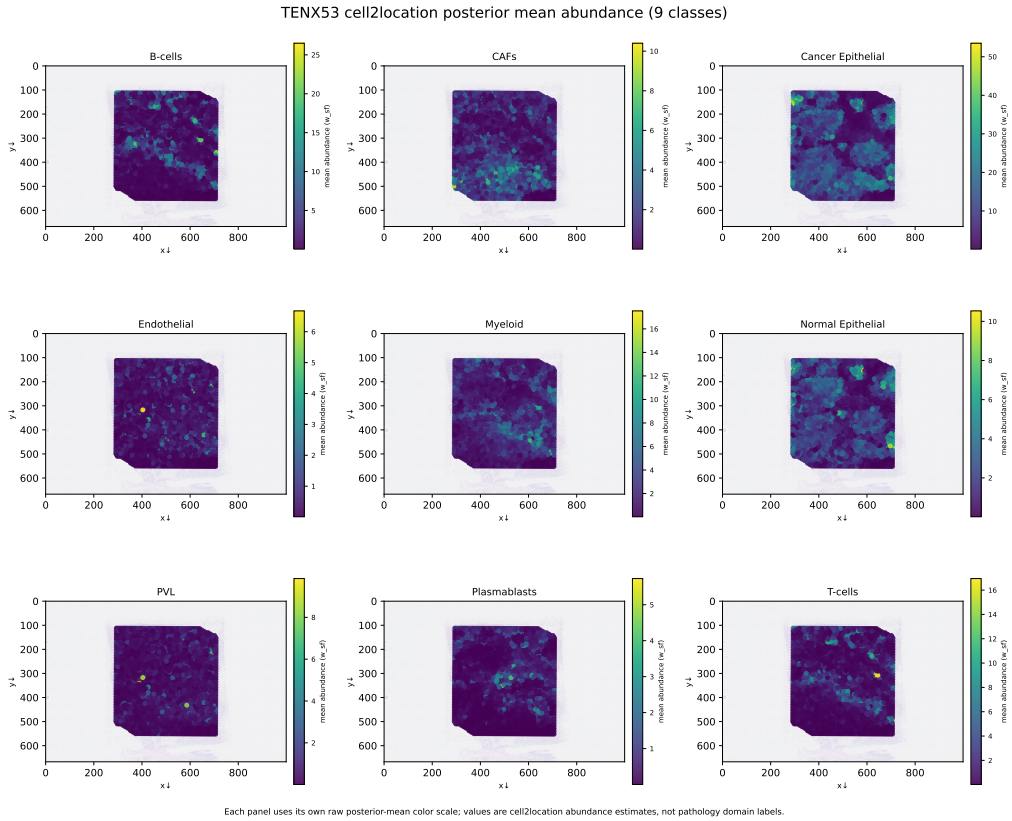


Figure S5: **Inferred cell-abundance context in TENX53.** Posterior mean cell2location abundances are shown at 4,898 aligned spots for B cells, CAFs, cancer epithelial cells, endothelial cells, myeloid cells, normal epithelial cells, PVL cells, plasmablasts, and T cells. The maps provide inferred cellular context, not manual pathology annotations or independent molecular validation (Kleshchevnikov et al., 2022).

using cancer-dependent molecular relationships, then used to select and combine image features before shared decoding. This design distinguishes gene identity from cancer-specific conditioning while retaining an image-derived expression readout.

### Q.3 Molecular priors and context-aware gene representations

Molecular priors describe genes from complementary perspectives. ESM-2 and RNA-FM encode protein and RNA sequences, while GenePT derives representations from functional descriptions (Lin et al., 2023; Chen et al., 2022; Chen & Zou, 2025). Comparisons of gene embeddings suggest that sequence-, expression-, network-, and text-based representations capture different aspects of gene function (Zhong et al., 2025). STRING adds general functional relationships that can complement these gene-level features (Szklarczyk et al., 2023).

Context-dependent resources provide another level of information. DepMap characterizes variation in genetic dependencies across cancer models (Pacini et al., 2024; Arafeh et al., 2025). Gene Ontology, Reactome, and Hallmark gene sets describe molecular functions and programs (The Gene Ontology Consortium, 2026; Milacic et al., 2024; Liberzon et al., 2015), while JUMP supplies perturbation-associated morphological profiles (Chandrasekaran et al., 2024). GeneRoute keeps these sources separate during adaptation, allowing their contributions to depend on the requested gene and available cancer context rather than treating all priors as interchangeable inputs.

## R BROADER IMPACT AND ETHICAL CONSIDERATIONS

Histology-based expression prediction may support exploratory molecular analyses of retrospective tissue collections. It may also reproduce biases associated with cohort composition, staining, scanners, institutions, and external knowledge sources. Outputs should remain clearly identified as predictions, with their uncertainty and provenance preserved. The model has not been evaluated for autonomous diagnosis or treatment selection and does not replace direct molecular measurements. Data and model reuse remains subject to source-specific licenses and applicable privacy, consent, and institutional requirements.

# S COMPLETE PERMANENT HELD-OUT GENE LIST

Table S5: **The 200 permanent held-out genes.** Versionless Ensembl identifiers are shown in the frozen registry order, read down each pair of columns. S denotes the stratum in Table A4; P denotes protein_coding, while N denotes genes with other GENCODE biotypes and RNA-FM. +/− indicates GenePT availability. The companion CSV also records gene types and source masks.

| Ensembl ID | S | Ensembl ID | S | Ensembl ID | S | Ensembl ID | S |
|---|---|---|---|---|---|---|---|
| ENSG00000006638 | P+ | ENSG00000137193 | P+ | ENSG00000168685 | P+ | ENSG00000198612 | P+ |
| ENSG00000027644 | P+ | ENSG00000137434 | P− | ENSG00000168887 | P− | ENSG00000203963 | P− |
| ENSG00000042304 | N− | ENSG00000137745 | P+ | ENSG00000169084 | P− | ENSG00000204387 | N+ |
| ENSG00000047188 | P+ | ENSG00000137841 | P+ | ENSG00000169314 | P− | ENSG00000204669 | P− |
| ENSG00000047315 | P− | ENSG00000138111 | P− | ENSG00000169609 | P− | ENSG00000204682 | N+ |
| ENSG00000059573 | P+ | ENSG00000138780 | P+ | ENSG00000170456 | P− | ENSG00000205045 | P− |
| ENSG00000072832 | P+ | ENSG00000139436 | P+ | ENSG00000171045 | P+ | ENSG00000205057 | N+ |
| ENSG00000074054 | P+ | ENSG00000140943 | P− | ENSG00000171067 | P− | ENSG00000205086 | N− |
| ENSG00000081803 | P+ | ENSG00000141012 | P+ | ENSG00000172244 | P− | ENSG00000205464 | N− |
| ENSG00000081853 | P− | ENSG00000141428 | P− | ENSG00000172466 | P+ | ENSG00000205710 | P− |
| ENSG00000082213 | P− | ENSG00000141433 | P+ | ENSG00000172728 | P− | ENSG00000211592 | N+ |
| ENSG00000088543 | P− | ENSG00000142784 | P+ | ENSG00000173077 | N+ | ENSG00000211675 | N+ |
| ENSG00000088888 | P+ | ENSG00000143110 | P− | ENSG00000174125 | P+ | ENSG00000211685 | N+ |
| ENSG00000093134 | N− | ENSG00000143157 | P+ | ENSG00000174370 | N− | ENSG00000211688 | N+ |
| ENSG00000094963 | P+ | ENSG00000143393 | P+ | ENSG00000174672 | P+ | ENSG00000211689 | N+ |
| ENSG00000100068 | N− | ENSG00000144061 | P+ | ENSG00000174899 | N− | ENSG00000211751 | N− |
| ENSG00000100802 | P− | ENSG00000144063 | P− | ENSG00000175262 | P− | ENSG00000211772 | N− |
| ENSG00000101003 | P+ | ENSG00000147894 | P− | ENSG00000175779 | N+ | ENSG00000211829 | N+ |
| ENSG00000101452 | P+ | ENSG00000148143 | P+ | ENSG00000176605 | N− | ENSG00000211890 | N+ |
| ENSG00000102302 | P+ | ENSG00000149609 | P− | ENSG00000176915 | P+ | ENSG00000211896 | N− |
| ENSG00000103089 | P+ | ENSG00000151136 | P− | ENSG00000176928 | P+ | ENSG00000211897 | N− |
| ENSG00000104067 | P+ | ENSG00000151292 | P+ | ENSG00000177946 | N− | ENSG00000211899 | N− |
| ENSG00000104412 | P+ | ENSG00000151470 | P− | ENSG00000179058 | P− | ENSG00000212719 | N+ |
| ENSG00000104972 | P+ | ENSG00000151694 | P+ | ENSG00000179088 | P− | ENSG00000213780 | P+ |
| ENSG00000105072 | P− | ENSG00000153113 | N− | ENSG00000179152 | P+ | ENSG00000214226 | P− |
| ENSG00000105427 | P+ | ENSG00000153283 | P+ | ENSG00000180999 | P− | ENSG00000221883 | N− |
| ENSG00000105501 | N− | ENSG00000153446 | P− | ENSG00000181215 | P− | ENSG00000222047 | N− |
| ENSG00000106105 | P− | ENSG00000154102 | P− | ENSG00000182919 | P− | ENSG00000225362 | N+ |
| ENSG00000106268 | P+ | ENSG00000155330 | P− | ENSG00000182957 | P− | ENSG00000225940 | N− |
| ENSG00000108448 | N− | ENSG00000156170 | P+ | ENSG00000182968 | P+ | ENSG00000227191 | N+ |
| ENSG00000111860 | P+ | ENSG00000157368 | P+ | ENSG00000183784 | N+ | ENSG00000228300 | P− |
| ENSG00000113231 | P− | ENSG00000157895 | P− | ENSG00000184530 | P− | ENSG00000228486 | P− |
| ENSG00000119559 | P− | ENSG00000158717 | P+ | ENSG00000185127 | P− | ENSG00000230510 | N− |
| ENSG00000119979 | P− | ENSG00000160179 | P+ | ENSG00000186501 | P+ | ENSG00000230626 | N− |
| ENSG00000121964 | P− | ENSG00000160218 | P+ | ENSG00000186654 | P+ | ENSG00000239521 | N− |
| ENSG00000123080 | P+ | ENSG00000161091 | P+ | ENSG00000187066 | P− | ENSG00000250486 | N+ |
| ENSG00000124333 | P− | ENSG00000161133 | N− | ENSG00000187699 | P− | ENSG00000253368 | P+ |
| ENSG00000125122 | N− | ENSG00000161980 | P+ | ENSG00000187944 | P− | ENSG00000254827 | N− |
| ENSG00000125462 | N+ | ENSG00000162598 | P− | ENSG00000187997 | P− | ENSG00000255221 | N− |
| ENSG00000125741 | P+ | ENSG00000164338 | P+ | ENSG00000188032 | P− | ENSG00000255552 | N− |
| ENSG00000128346 | P− | ENSG00000164393 | N− | ENSG00000188086 | N− | ENSG00000260456 | P− |
| ENSG00000132514 | P+ | ENSG00000164746 | P− | ENSG00000188707 | N− | ENSG00000261587 | P− |
| ENSG00000133026 | P+ | ENSG00000165164 | P+ | ENSG00000189366 | N− | ENSG00000262484 | P+ |
| ENSG00000133475 | N− | ENSG00000165182 | P− | ENSG00000196381 | N− | ENSG00000267221 | P− |
| ENSG00000134013 | P+ | ENSG00000165526 | P+ | ENSG00000196850 | P+ | ENSG00000270806 | P− |
| ENSG00000135045 | P− | ENSG00000166323 | P− | ENSG00000197364 | N− | ENSG00000275464 | N− |
| ENSG00000135502 | N− | ENSG00000166845 | P− | ENSG00000197558 | N+ | ENSG00000277734 | N+ |
| ENSG00000135951 | P+ | ENSG00000167447 | P+ | ENSG00000197568 | N− | ENSG00000278505 | P− |
| ENSG00000135974 | P− | ENSG00000167747 | N− | ENSG00000197808 | P+ | ENSG00000283597 | N− |
| ENSG00000136758 | P+ | ENSG00000167920 | N− | ENSG00000198019 | N− | ENSG00000284308 | P− |